DOCUMENT RESUME

| | | |
|---|---|---|
| ED 328 773 | | CE 056 988 |

| | |
|---|---|
| AUTHOR | Friedlander, Daniel; Long, David |
| TITLE | A Study of Performance Measures and Subgroup Impacts in Three Welfare Employment Programs. Research Report Series RR-87-28. |
| INSTITUTION | National Commission for Employment Policy (DOL), Washington, D.C. |
| SPONS AGENCY | Family Support Administration (DHHS), Washington, DC. Office of Family Assistance.; Office of the Assistant Secretary for Planning and Evaluation (DHHS), Washington, D.C. |
| PUB DATE | Mar 87 |
| NOTE | 110p. |
| PUB TYPE | Reports - Research/Technical (143) |
| EDRS PRICE | MF01/PC05 Plus Postage. |
| DESCRIPTORS | Adult Education; Employment Patterns; *Employment Programs; *Evaluation Criteria; Federal Programs; *Females; Followup Studies; Income; Predictor Variables; *Program Effectiveness; Vocational Followup; *Welfare Recipients; *Welfare Services |
| IDENTIFIERS | *Aid to Families with Dependent Children; California (San Diego); Maryland (Baltimore); Virginia |


ABSTRACT

This study (the first part of a two-part study) analyzed the effectiveness of three mandatory welfare employment programs in serving different segments of the Aid to Families with Dependent Children (AFDC) caseload. Data were collected in evaluations of welfare employment initiatives in San Diego, Baltimore, and several counties in Virginia. Program participation was required for different portions of the AFDC caseload. Eligible applicants and recipients (primarily female) were randomly assigned to experimental groups, which received program services, or to control groups, which did not. Data were collected using AFDC payments and Unemployment Insurance earnings records. With few exceptions, employment and earnings impacts were consistently smaller than average for the welfare applicants and recipients who had the best work records and the least prior welfare experience. The impacts were usually larger for more dependent individuals, although not for the cases that were the most dependent. Programs had less consistent impacts on subgroups defined by characteristics such as marital status and educational level. The outcome measures examined were not valid indicators of program performance. Neither job entries nor cases off welfare were a satisfactory predictor of the changes in employment, earnings, and welfare receipt achieved by the programs studied. (20 references) (YLB)

# A STUDY OF PERFORMANCE MEASURES AND SUBGROUP IMPACTS IN THREE WELFARE EMPLOYMENT PROGRAMS

by

Daniel Friedlander
David Long
MDRC

March 1987 RR-87-28

RESEARCH REPORT SERIES
NATIONAL COMMISSION
FOR EMPLOYMENT POLICY
1522 K STREET, N.W.
WASHINGTON, D.C. 20005

This analysis completes the first phase of ongoing research being conducted on performance measurement and subgroup impacts in welfare employment programs. This first phase was funded by the Office of the Assistant Secretary for Planning and Evaluation at the U.S. Department of Health and Human Services; by the Office of Family Assistance, Family Support Administration, also part of the U.S. Department of Health and Human Services; and by the National Commission for Employment Policy. The findings and conclusions of this report do not necessarily represent the official positions or policies of the funders.

The conclusions and recommendations in this report are those of the contractor and do not necessarily reflect the views of the National Commission for Employment Policy.

## ACKNOWLEDGMENTS

The authors wish to thank those at the U.S. Department of Health and Human Services (the Office of Assistant Secretary for Planning and Evaluation and the Office of Family Assistance) as well as staff at the National Commission for Employment Policy who made this research possible. Some staff members from these agencies also participated in an early review of the findings and provided thoughtful comments throughout the analysis. In particular, we are grateful for the valuable input and suggestions from Deirdre Duzor, Ken Lee, M.I. Pendell, Canta Pian, Howard Rolston, Steve Sandell, and Daniel Weinberg. The study benefited greatly from the suggestions and advice of Gary Burtless at the Brookings Institution, David Ellwood of Harvard University, and Frank Levy, of the University of Maryland. Gratitude is also expressed to the Work/Welfare Committee of MDRC's Board of Directors and to Gordon Berlin of the Ford Foundation who reviewed and commented on the results of this research as did the special Advisory Committee to the Work/Welfare Demonstration.

Because this study is based on data collected in MDRC's ongoing Demonstration of State Work/Welfare Initiatives, the authors also express gratitude to many members of MDRC staff who have obtained the data and produced the state reports over the last few years. Others on MDRC staff were instrumental in shaping this report -- in particular, Judith Gueron, Michael Bangser, Michael Borus, Barbara Goldman and John Wallace.

Additionally, at MDRC, Marjorie Erickson assisted in the review of the literature and oversaw preparation of the data sets and the initial computer analysis, a task taken over by Jan Bryant. Ginger Knox and Emma Caspar worked on benefit-cost calculations. David Jaeger, Stephanie Powell and Rudd Kierstead provided programming support. Sheila Mandel edited this report, while Naomi Weinstein prepared the tables.

## EXECUTIVE SUMMARY

This report presents a preliminary analysis of the effectiveness of three mandatory welfare employment programs in serving different segments of the Aid to Families with Dependent Children (AFDC) caseload. The analysis, covering the first phase of a two-part study, has been undertaken to obtain two kinds of information that are useful in designing and operating such programs. One is estimates of the programs' relative impacts on the employment and welfare receipt of different groups of welfare recipients. These impacts may indicate groups to which program services can best be targeted in order to use funds efficiently. The other is the development and validation of short-term performance indicators, which are important in judging these programs' performance in meeting their long-term objectives of increasing employment and reducing welfare dependency.

The analysis is based on data collected in evaluations of welfare employment initiatives in San Diego, Baltimore and several counties in Virginia. Participation in the programs was required for different portions of the AFDC caseload who are "mandatory" under federal Work Incentive (WIN) Program regulations. The programs also provided different services and operated in different labor markets.

The populations served and the three programs' services are as follows. The San Diego program required the participation of new AFDC applicants in a three-week job search workshop. Those who did not find jobs during this time were then assigned to a 13-week work experience

position in a public or nonprofit agency. Baltimore's program required both applicants and newly-mandatory recipients to participate, but activities could be selected from a number of job search, work experience, education and training options. In Virginia, the program required job search of the entire WIN-mandatory caseload, which was sometimes followed by work experience, education or training. This program differed from the others in that it operated in rural as well as urban areas of the state.

Each of the three evaluations used experimental research designs to estimate program impacts. Eligible applicants and recipients were randomly assigned to experimental groups, which received program services, or to control groups, which did not. (It should be noted that an applicant for welfare at the time of random assignment was called an "applicant" throughout the study, even though many became recipients.) The experience of the control groups -- which could have received services from sources other than the programs -- indicates what would have happened to the experimental groups in the absence of the programs, providing a benchmark against which to measure program impacts.

Data were collected using AFDC payments and Unemployment Insurance earnings records for varying periods of up to three years in San Diego and Baltimore; only a short follow-up period was available in Virginia. The data considered in this analysis are for single-parent (primarily female) heads of households. Two-parent households (primarily men eligible under the AFDC-Unemployed Parent program) were included in two of the program evaluations, but are not part of this study's sample.

The distinction between "outcomes" and "impacts" underlies most of the findings of this analysis. An "outcome" is the employment or welfare

status of a person at a specified point after program enrollment. An "impact" is the change in outcomes produced by a program during that period, or simply the outcome difference between the experimental and control groups. Program impacts are smaller than outcomes because the normal job-finding and welfare departure rates of the AFDC population -- i.e., the control group's level -- are not zero in the absence of a program. Past research, however, has indicated that groups exhibiting worse-than-average outcomes may generate better-than-average impacts.

Subgroup Impact Differences

The analysis focuses on female WIN-mandatory AFDC subgroups defined by two characteristics: prior work and welfare history. The samples were divided into subgroup categories according to simple objective measures of job-readiness and welfare dependence at the time these individuals became eligible for the program (i.e., were randomly assigned). Three subgroups were based on earnings from employment in the year prior to random assignment: no earnings, $1 to $2,999, or $3,000 or more. Similarly, three other subgroups were created according to the length of time that these people had been on welfare (that is, had had their own AFDC case) before random assignment: never, two years or less, or more than two years.

Other characteristics, such as marital status and prior education, were also examined, but their role in determining impacts was not as consistent across programs as the prior earnings and welfare dependency measures.

- When subgroups we defined by previous work and welfare experience, the mo job-ready and least welfare-dependent groups had below-average program impacts, which were often the smallest impacts.

With few exceptions, employment and earnings impacts were consistently smaller than average for the welfare applicants and recipients who had the best work records and the least prior welfare experience. Frequently, program impacts on these groups were the smallest. This does not mean that the more job-ready and less dependent people who enrolled in these programs were less able to find jobs than those with poor records. In fact, as one would expect, these people entered employment much more frequently. But control group members with the better work records or no welfare experience also found employment almost as easily, so program interventions made less of a difference with these groups.

This point is demonstrated in Table 1, which shows composite estimates from the separate samples of AFDC applicants and recipients in the three programs analyzed. These estimates should be interpreted with care because they do not show the underlying variation across programs. Nevertheless, the table indicates that xperimentals with $3,000 or more in earnings in the pre-program year -- the highest earnings category -- achieved an average employment rate of 62 percent per quarter. At the same time, experimentals who had not worked at all in the year prior to program enrollment had only a 26 percent employment rate. Yet the employment impact for the first group was somewhat below the average, at 3.1 percentage points. The "less employable" group attained a 4.9 percentage point gain, the highest of the three prior-earnings categories.

Similarly, individuals who had never had an AFDC case in the past

TABLE 1

COMPOSITE PROGRAM IMPACTS ON EMPLOYMENT AND EARNINGS
OF AFDC APPLICANTS AND RECIPIENTS, BY MAJOR SUBGROUP

| | Quarterly Employment Rate Quarters 4 - Last (%) | | | |
|---|---|---|---|---|
| Subgroup | Experimental | Control | Difference | Percent Improvement |
| Full Sample | 38.0 | 33.9 | + 4.1*** | +12 |
| Prior Year Earnings | | | | |
| $3000 or More | 62.0 | 58.9 | + 3.1 | + 5 |
| $1-2999 | 48.3 | 45.0 | + 3.3* | + 7 |
| None | 26.0 | 21.1 | + 4.9*** | +23 |
| Had Own AFDC Case | | | | |
| Never | 42.1 | 42.4 | - 0.3 | - 1 |
| Two Years or Less | 43.5 | 38.2 | + 5.2*** | +14 |
| More Than Two Years | 33.8 | 29.0 | + 4.8*** | +17 |

| | Average Earnings Per Quarter Quarters 4 - Last ($) | | | |
|---|---|---|---|---|
| Subgroup | Experimental | Control | Difference | Percent Improvement |
| Full Sample | 638 | 550 | + 87*** | +16 |
| Prior Year Earnings | | | | |
| $3000 or More | 1323 | 1235 | + 88 | + 7 |
| $1-2999 | 747 | 699 | + 48 | + 7 |
| None | 397 | 291 | +107*** | +37 |
| Had Own AFDC Case | | | | |
| Never | 828 | 811 | + 17 | + 2 |
| Two Years or Less | 788 | 661 | +127*** | +19 |
| More Than Two Years | 494 | 406 | + 88*** | +22 |

SOURCE: MDRC calculations from California, Maryland, and Virginia Unemployment Insurance earnings records and program information systems.

NOTES: "Quarterly Employment Rate" is the percent of experimentals or controls ever employed in a quarter; "Average Earnings Per Quarter" is the average earnings of experimentals or controls in a quarter, whether or not they were ever employed; and "Percent Improvement" is the experimental-control difference expressed as a percent of the mean for controls. These composite estimates are weighted averages of the regression-adjusted estimates for AFDC applicants and recipients in the San Diego, Baltimore and Virginia programs.

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.

achieved above-average employment rates, but experienced virtually no impacts in employment and earnings, while those with extensive prior welfare experience had lower employment rates but showed larger employment gains.

Program impacts on welfare incidence and the amount of welfare payments were smaller than impacts on employment and earnings. Composite estimates are shown in Table 2. Again, sample members with high earnings and low prior welfare receipt often showed relatively smaller welfare impacts, although the overall pattern was less consistent than for employment impacts.

- The impacts were usually larger for more dependent individuals, although not for the cases that were most dependent. This suggests that some program models may operate most effectively with individuals above some threshold level of employability.

While the impacts of the three programs on employment and welfare were often larger for the more dependent segments of the AFDC caseload, this was not uniformly true. For example, the impacts for recipients in both Baltimore and Virginia -- who, by definition, had been on welfare for a period of time -- were substantially smaller than for applicants. In fact, the applicant impact on quarterly earnings was about three times the size of the recipient impact.

These findings suggest that the relationship between individual dependency and program impacts is not linear. In Figure 1, estimates of the San Diego and Baltimore program impacts on earnings were plotted against an estimated dependency score for each individual (reflecting predicted welfare use and earnings) based on prior work, welfare experience

TABLE 2

COMPOSITE PROGRAM IMPACTS ON WELFARE RECEIPT
OF AFDC APPLICANTS AND RECIPIENTS, BY MAJOR SUBGROUP

| Subgroup | Monthly AFDC Receipt Rate Quarters 4 - Last (%) | | | |
|---|---|---|---|---|
| | Experimental | Control | Difference | Percent Improvement |
| Full Sample | 47.6 | 48.9 | - 1.3 | - 3 |
| Prior Year Earnings | | | | |
| $3000 or More | 33.9 | 34.2 | - 0.3 | - 1 |
| $1-2999 | 44.2 | 45.0 | - 0.8 | - 2 |
| None | 52.5 | 54.7 | - 2.2* | - 4 |
| Had Own AFDC Case | | | | |
| Never | 30.9 | 33.1 | - 2.2 | - 7 |
| Two Years or Less | 40.6 | 41.6 | - 1.0 | - 2 |
| More Than Two Years | 57.4 | 58.8 | - 1.4 | - 2 |

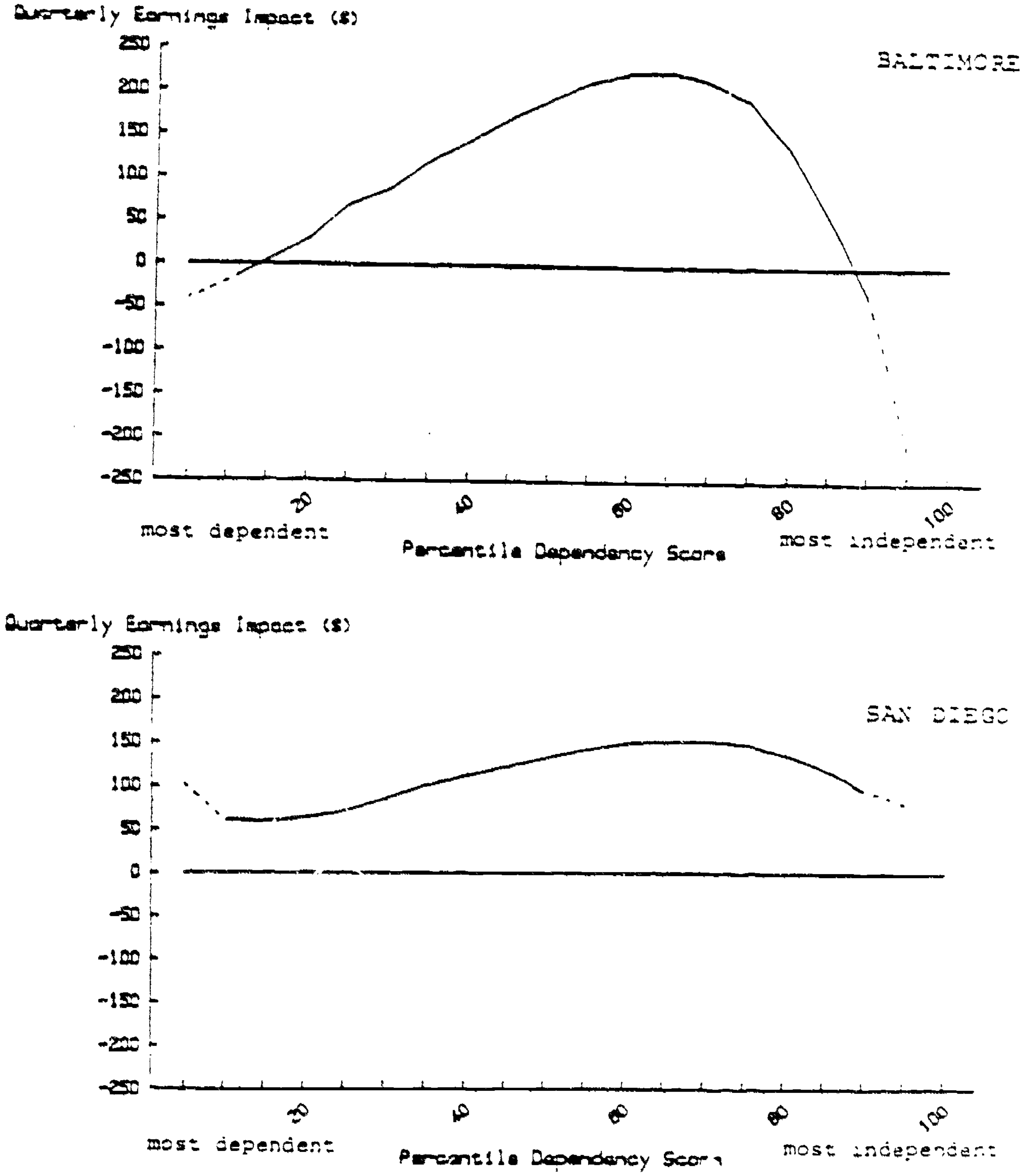
| Subgroup | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | | |
|---|---|---|---|---|
| | Experimental | Control | Difference | Percent Improvement |
| Full Sample | 425 | 441 | - 16* | - 4 |
| Prior Year Earnings | | | | |
| $3000 or More | 301 | 309 | - 7 | - 2 |
| $1-2999 | 388 | 405 | - 18 | - 4 |
| None | 474 | 495 | - 21* | - 4 |
| Had Own AFDC Case | | | | |
| Never | 297 | 328 | - 31 | - 9 |
| Two Years or Less | 361 | 373 | - 12 | - 3 |
| More Than Two Years | 503 | 519 | - 17 | - 3 |

SOURCE: MDRC calculations from California, Maryland and Virginia welfare records and program information systems, and in Virginia, from Fairfax County AFDC case files.

NOTES: "Monthly AFDC Receipt Rate" is the percent of experimentals or controls who received AFDC payments during a month; "Average AFDC Payments Per Quarter" is the average amount of payments received by experimentals or controls over a quarter, whether or not they received any; and "Percent Improvement" is the experimental-control difference expressed as a percent of the mean for controls. These composites are weighted averages of the regression-adjusted estimates for AFDC applicants and recipients in the San Diego, Baltimore and Virginia programs.

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.

FIGURE 1
BALTIMORE AND SAN DIEGO EXPERIMENTALS:
QUARTERLY EARNINGS IMPACTS, BY DEPENDENCY SCORE




NOTE: Dependency scores were estimated for San Diego and Baltimore experimentals based on their predicted earnings and predicted welfare receipt. The score is stated in percentile form, with "0" representing the most dependent. A score of 20 indicates that 20 percent of the sample ranked more dependent. Segments near endpoints of the curve are estimated with less precision and are therefore indicated with dashes.

and other characteristics, such as number of children and educational attainment.

In Baltimore -- where the program worked with cases over a wide range of dependency -- clients with the highest predicted welfare receipt and lowest earnings (the far left-hand side of Figure 1) did not appear to benefit from the particular services offered. Above some threshold level of dependency, the impact on earnings increased. But at the other end of the spectrum (the far right-hand side of the figure), the program again had less effect. The relatively job-ready WIN-mandatory caseload seemed better able to enter employment and leave welfare without program help. Thus, this program had its greatest effect on the large block of enrollees in the middle.

San Diego served a less dependent population -- only the AFDC applicants. Again, impacts are smaller for those at either end of the dependency spectrum (see the bottom graph in Figure 1), although San Diego's curve is less pronounced than Baltimore's one.

Dependency impact profiles should be developed for more than just two programs before final conclusions are drawn. Different service models may produce different profiles. For example, program services planned especially for highly dependent individuals (such as supported work) or for relatively job-ready individuals (such as job placement assistance) may have very different contours. It may also be important to consider program performance in relationship to the different labor markets. Judging from the limited results available thus far from rural counties in Virginia -- where the economic conditions were very different from those in San Diego in Baltimore -- program experience did not appear to fit the pattern of

Figure 1.

- The programs had less consistent impacts on subgroups of the WIN-mandatory AFDC caseload who were defined by characteristics such as marital status and educational level.

While many factors together may contribute to the impact results, single characteristics other than prior earnings and welfare history did not generally produce consistent impact differences in this study. For example, the number of children in a household, which is related to welfare dependency, is not alone a consistent explanatory variable for impacts. This was also true for variables such as race and marital status. Prior employment and welfare receipt are thus the most important characteristics to consider when trying to improve the results of welfare employment programs for WIN-mandatory individuals.

However, some of the other characteristics were important in specific program settings. For example, a higher level of education was positively related to impacts in the San Diego program, which did not offer educational services and was designed to move people into the labor market quickly. That was not true in Baltimore, which did offer remedial education services.

## Program Performance Measures

While program performance should be ideally assessed in terms of impacts, only short-term outcome measures such as "job entries" (placements) and cases "off-welfare" (case closures) are available in most instances. This subgroup impact analysis suggests not only that these measures overstate impacts, but that they also misrepresent the relative

performance of certain subgroups of welfare recipients. Thus, unless subgroup differences are taken into account, current performance measures may be sending the wrong signals to program administrators about the groups who should be receiving priority for program services. It is important to note, however, that other measures -- such as average entry wage levels -- could not be addressed in this analysis.

- Unadjusted "job entry" and "off-welfare" measures were poorly correlated with the employment and welfare impacts of the programs in San Diego and Baltimore. Hence, these measures by themselves are not good indicators of program performance.

The relationship between outcomes and program impacts was examined by estimating impacts for each member of the experimental group and then determining the correlation of these estimates with the outcome measures. The conclusion is that the outcome measures examined were not valid indicators of impacts. Neither job entries nor cases off-welfare were a satisfactory predictor of the changes in employment, earnings and welfare receipt achieved by the programs studied. These findings remained true when differential program costs were considered.

This conclusion -- which runs counter to common wisdom -- simply reflects the fact that the magnitude of the program effect on finding a job or leaving welfare is greater for some groups of individuals than others. This does not imply that programs should stop trying to help all people in the caseload find jobs and leave welfare. It does mean that judging programs on the basis of these outcome measures -- without considering differences in caseload characteristics and economic conditions -- is unwise. It is quite possible, for example, for a program with a relatively low placement rate in a poor labor market to have greater impacts than

another program with a more job-ready caseload and more placements. The analysis also shows that this conclusion does not change when longer-term employment rates are substituted for immediate job entries.

- Weighting performance measures to reflect prior work histories greatly improved their value in predicting program impacts.

Giving more weight to job entries and movement off welfare by cases with no or limited employment experience improved the correlation between the performance measures and earnings impacts. One weighting scheme tested gave four points for a job entry by a person not employed in the previous year and two points or one point to people who had some pre-program earnings -- $1 to $2,999 or $3,000 or more. A number of other weighting procedures were tried, and some of these were also an improvement over unweighted indicators.

- Like job entries and welfare case closures, simple program participation measures can give a misleading impression of program performance. Weighted participation or "program coverage" measures -- while more difficult for program operators to use -- may be better suited to assessing the performance of mandatory welfare employment programs.

Performance measures based on participation -- that is, active enrollment in program services or activities -- are sometimes used in addition to job entry and welfare outcome measures. Participation standards can be important because they have the advantage of encouraging program operators to serve a broad range of those eligible. However, these measures also have some drawbacks, especially for mandatory welfare employment programs, which have sanctions that reduce welfare grants for individuals who do not cooperate with participation requirements. These programs intentionally attempt to affect the behavior of nonparticipants as well as participants.

Moreover, because this analysis suggests that unweighted participation measures may misrepresent any program effectiveness that is related to participation, priorities or weighting schemes for AFDC subgroups should be considered if these measures are to be used.

Program "coverage" measures provide a possible alternative, although so far they have only been used as an analytical tool in program evaluation. In measuring the number of people covered by a program, a broader view of program contact is taken than just participation. The number of cases in which participation is no longer required -- because someone becomes employed or leaves AFDC on his or her own -- as well as those in which sanctions for nonparticipation have been imposed, are counted in addition to cases of participation. The proportion of "uncovered" cases directs attention to the group the program has not reached -- that is, the individuals who are still on welfare, unemployed, have not begun to satisfy program requirements, or have not been sanctioned for noncompliance.

Such measures, however, are generally not used at present and have a number of practical limitations, including extensive and potentially expensive changes in data collection procedures.

## Conclusions and Open Issues

The research reported in this document addresses a number of important issues in the monitoring and targeting of welfare employment programs. It also raises questions relevant to the broader employment and training delivery system. While the results to date are striking and suggest the promise of further research, they should be considered preliminary and suggestive, rather than definitive. In some cases, the implications are

quite clear. But in others, they raise questions to which the appropriate policy response is less clear.

For example, a convincing lesson from this study is that, if resources are limited, it is a mistake to concentrate only on serving the most job-ready portion of the AFDC caseload. Since this was the tendency in the WIN program, this is an important message and suggests a shift in strategy. Thus, performance measures should be revised to encourage programs to work with more dependent and less job-ready individuals. Unweighted outcome measures clearly do not do this. Weighted measures create more appropriate incentives by explicitly taking subgroup differences into account. Many are already recognizing this lesson and adopting measures to try to adjust service priorities and monitoring tools.

On the other side, readers should be cautioned that the results do not yet suggest an exclusive focus on the more disadvantaged or the immediate adoption of one particular weighting scheme. These cautions are suggested by several factors. First, while impacts were smaller for the more job-ready, they were sometimes positive. Second, and more important, the results reported in this analysis were for programs that made no targeting choices and thus mixed in job clubs, placement efforts and other activities for individuals with a wide range of prior work experience and other "employability" factors. This study thus cannot say whether similar positive impacts for the more disadvantaged could be obtained by programs that served only such groups.

One could well imagine, for example, that including the more job-ready in job search workshops helped motivate both program staff and the most disadvantaged and thus contributed to the positive results reported here.

This "mainstreaming" hypothesis is not tested in this study, but it suggests that administrators should look carefully at the operational results of more targeted services before exclusively using resources for this group. In addition, working only with individuals with lower skills and measured outcomes could have political, administrative or stigmatizing effects. For example, it may be difficult to convince people that a placement rate of 30 percent represents a substantial positive achievement. Such low rates may also discourage staff efforts. And, employers may think differently about a work program that refers only clients with no prior work history.

The results of this study are most convincing when they suggest not serving only the most job-ready but rather serving a broad range of the caseload, with differential rewards or monitoring structures. They do not yet confirm exclusive targeting.

Finally, the weighting schemes examined in this analysis were only tested in two programs. It will be important to see whether their advantages hold up with different groups and programs in different states before a particular formula is adopted. Thus, while the results to date indicate possible directions to go in improving program monitoring, they do not prescribe a formula that would be valid in a wide range of economic, demographic and programmatic conditions.

In addition, readers should be aware of a number of caveats and open questions. First, the results presented here come from mandatory programs enrolling everyone within a specified group of welfare recipients. Very different issues and lessons could arise in selective programs that can choose the people they wish to enroll. Program operators, for example,

19

could screen intensely among the more disadvantaged, possibly identifying only the most motivated within this group, and thus undercut the very message implicit in the results reported here.

A major open question arises from the preliminary finding that -- at least for the relatively inexpensive and often non-intensive services studied -- there may be a threshold effect: i.e., impacts may be smaller for the most dependent persons. It will be important to examine whether this is also true for programs that provide more intensive services. Notably, can programs offering intensive educational remediation or long-term education and skills training change the shape of the impact curve in Figure 1 and succeed in increasing the earnings of the most disadvantaged? Results from another study -- Supported Work -- suggest that at least that treatment had substantial benefits for some members of this group.

Finally, performance measures are only useful if they can be implemented: the data must be available and the calculations possible to do in a reasonable period of time. The analysis in this report drew on an unusual data base, which is not readily available to program administrators. It will be important to examine the feasibility and cost of adopting some of these measures.

Some of these questions require further operational experience, and some go beyond what can be learned from the programs included in MDRC's study. Others will be addressed during the second phase of the planned research, drawing on the large knowledge base of this study and on the promising directions seen so far.

TABLE OF CONTENTS



21

## LIST OF TABLES

## LIST OF FIGURES

A STUDY OF PERFORMANCE MEASURES AND
SUBGROUP IMPACTS IN THREE
WELFARE EMPLOYMENT PROGRAMS

CHAPTER 1

# INTRODUCTION

The search for valid and workable standards of performance to be used in employment programs for welfare recipients has been one of the major themes in current efforts to reform welfare policy. Such close attention is warranted because performance measures are one of the primary means by which broad policy is translated into the specific objectives that guide the operations of programs. Standards allow administrators to assess how well their programs are doing, to evaluate the worth of innovative programs, and to identify problems in existing models. They can also influence the programs' service priorities, encouraging a focus on the welfare groups most likely to help the programs achieve a high performance rating. In this manner, standards also influence the allocation of funds and, in a period of fiscal restraint, it is important that performance measures promote efficient utilization of resources.

Given this importance, performance measures should be appropriate for the programs that use them. Poorly designed or inadequately tested performance standards can work against the objectives of the authorizing legislation. They can waste staff time and other program resources, with the result that neither the welfare population nor society is well served.

This paper examines performance monitoring by studying three employment and training programs for recipients of Aid to Families with Dependent Children (AFDC) -- programs in which participation was mandatory. It is the first phase of a two-part investigation into the differences in

the impacts of such programs on the employment and welfare receipt of selected AFDC subgroups. The study uses data from the Demonstration of State Work/Welfare Initiatives, a five-year, eight-state series of large-scale social experiments conducted by the Manpower Demonstration Research Corporation (MDRC). The data are unusual in that the research samples they describe were generated in controlled experiments involving random assignment. They are also comprehensive enough to permit program performance to be considered in terms of multiple program effects as well as program costs.[1] Complete subgroup analyses are presented for two programs, with a preliminary analysis offered for the third program, which has limited follow-up data at this time.

It should be emphasized that all of the programs were targeted to AFDC case heads meeting the Work Incentive (WIN) program definition of mandatory: single parents (mostly women) who had no child under the age of six, and had no other known barriers to participation. This so-called WIN-mandatory group makes up about one-third of the AFDC caseload nationwide. Unemployed heads of two-parent households, who are also mandatory, were part of the research in the states that served this group, but these samples have been excluded because they are primarily men, with different work backgrounds, and receive assistance under different rules.

This analysis uses the subgroup impacts generated from the experimental data on the three programs to evaluate the validity of two frequently used performance measures: the number of "job entries" (placements) and the number of cases "off-welfare" (case closures).[2] Some alternative standards are also considered. However, the implications of the analysis are somewhat broader in scope than welfare employment programs because many of

the issues examined are common to other programs for low-income or disadvantaged groups, such as those funded by the Job Training Partnership Act (JTPA). This study's focus on testing employment and welfare receipt measures should not imply that measures such as wage rates, job retention and participation have no validity for some programs.

The discussion is structured as follows. This chapter reviews issues relevant to welfare population subgroups and program performance. Chapter 2 discusses the welfare employment programs studied and their research designs, and is followed in Chapter 3 by an explanation of the methodology used to estimate subgroup impacts and to test performance indicators. Chapters 4 and 5 are central to the analysis. Chapter 4 presents impacts and costs for the major subgroups in the study, while Chapter 5 evaluates the validity of alternative performance measures, using program impact estimates. Since this is the first phase of the study, no conclusions are as yet offered.

## A. Issues in Assessing Program Performance Monitoring

Performance measures are intended to promote program effectiveness, conserve resources, and ensure compliance with overall goals and directives. A wide range of indicators has been developed and used in the WIN program, those funded by the Comprehensive Employment and Training Act (CETA), and, more recently, by the Job Training Partnership Act. Historically, job placements and welfare reductions have been the most important indicators in WIN. These measures have seemed useful in conveying program achievements in straightforward terms to policymakers and the general public. Their incorporation into the fiscal WIN Allocation Formula under-

lined their significance to operators of welfare employment programs. Other indicators, however, were also part of the WIN Allocation Formula, such as the quality of job entries, usually measured by wage rates and job retention.[3] In measuring employment outcomes, all enrollees have been counted with equal weight.

These indicators all measure the outcomes of a registrant's program experience at some point _after_ registration. Another set of indicators looks at the activity of registrants while _in_ the program; these include counts of registrants, participants, program completors and similar measures. Participation data have been examined in evaluations of WIN, CETA and other programs, but the trend today has been to deemphasize these indicators, even though they provide immediate feedback and are relatively inexpensive to collect.[4] Instead, emphasis has been on measures that communicate program goals in terms of post-program outcomes. For example, the JTPA legislation explicitly requires that standards for adult participants be based on job entries, wages and earnings, retention and welfare reductions.[5]

### 1. Outcomes and Impacts

The distinction between "outcomes" and "impacts" is critical to an understanding of how well outcomes measure program performance. An outcome is the employment and/or welfare status of a person at some point in time after program registration. Hence, the outcome "employed and not receiving welfare at quarter 4" describes the status of a person 9 to 12 months after program entry.

The real effects of a program cannot be judged by outcomes, however, given the high degree of normal job-finding and welfare departure within

the welfare population. Program impacts, in contrast, do state the true program effects -- if they have been correctly estimated. Impacts measure a change in behavior, one that can be estimated by comparing the behavior of a group of people who receive the program treatment with that of a similar group of people who do not receive the treatment: i.e., a control group, the behavior of which in the three programs studied is discussed in Chapter 3. The distinction between a level -- the outcome -- and a change -- the impact -- is important because program impacts are likely to be far smaller than program outcomes, since factors such as the control group's employment rate are not zero in the absence of a program.

Past research has suggested that groups exhibiting worse-than-average outcomes may, in fact, experience better-than-average program impacts. For example, an evaluation of a job search and work experience program operated in San Diego found that 73 percent of WIN-mandatory AFDC applicants who had worked at some time during the year prior to their program entry were able to find employment during the year and one-half following enrollment. This high rate was, in fact, only a 2 percentage point change from the control group employment level -- that is, the rate that applicants with a prior work history were able to achieve on their own. In contrast, enrollees without prior employment attained only a 48 percent employment rate, but this outcome was a 10 percentage point increase, or impact, from the control group's employment rate of 38 percent.[6]

Given these patterns, performance indicators based only on outcomes create a misleading impression of program effectiveness. Clearly, they overstate program impacts because the measures have no comparison against which to judge change. However, a problem more serious than simple

overstatement may exist. Program resources may be ineffectively targeted if these standards place emphasis on serving the least appropriate groups -- that is, those who would have done well on their own, without the programs. Conversely, people who could benefit most from these programs may be underserved. The important role of performance measures in determining how programs are operated and how resources are allocated is the principal reason that this examination of current performance measures has been undertaken.

The findings in this paper and similar ones from other studies suggest that consideration be given to the development of performance formulas that do not treat each person's outcome equally. Such formulas allow outcome standards to vary by local economic conditions, registrant characteristics, and even by service components. Regression adjustment is one way to develop formulas that permit more flexible performance standards for programs serving groups with a low likelihood of finding employment readily, or those operating in areas with relatively poor labor markets, where it is hard to find jobs. In such plans, performance weights are based on many background variables, such as prior work experience, the length of welfare dependency, education and number of children.

Multiple regression formulas have advantages, but they can be complex. They may also be more suited to analysis at the aggregate level than for the communication of program objectives to local staff or in setting performance criteria for service subcontractors. Moreover, the correct regression weights may not have been used in the past, many having been based on outcome levels rather than estimated impacts.

This study presents some simpler weighting options, which take one, or

perhaps two, characteristics into account instead of many. Prior employment is one important characteristic for WIN-mandatory AFDC women, as this report will show. However, this study -- while searching for better ways to judge program success -- makes no pretense of having all of the answers, for the goals of some programs may not be easily translated into simple weighting schemes.

2. <u>Issues in Targeting</u>

Much of the recent work in targeting welfare employment programs has focused on AFDC subgroups outside the WIN-mandatory category -- such as mothers with young children, who are not part of this study. This ongoing research has tried to identify subgroups to whom services should be targeted because they are likely to have relatively long periods of welfare dependency.[7] The basic premise is that the longer the predicted period of dependency, the greater the potential reduction in dependency that program services can produce. A key assumption is that treatments can be found that would work effectively with the most dependent subgroups.

These studies have successfully linked differences in length of welfare dependency with measured recipient characteristics. One important finding has been that the majority of people who enter the welfare system spend less than four years on the rolls, even counting repeat spells. Services targeted to this group, it is argued, may not be an efficient use of resources. The smaller proportion of people who remain on welfare account for the bulk of AFDC benefit expenditures, with one study estimating that as much as 60 percent of all grant outlays are paid to only 25 percent of all recipients.[8] Program assistance targeted to these recipients, it is claimed, may substantially decrease the costs of

dependency, again assuming that effective services can be found for this group.

A study by David Ellwood maintains it is possible to identify, on the basis of demographic characteristics, the subgroups with a high risk of extended dependency. His analysis identifies young, never-married women, as well as women with young children, as candidates for targeting. As an alternative, he would let the most dependent identify themselves -- that is, those remaining on welfare after some specified period of time would receive program services.[9]

The conditions and problems that lead to extended dependency, however, may not be amenable to change with low-cost employability services.[10] This question cannot be resolved in this study. Further, the subgroups many researchers identify as the portion of the AFDC caseload with the longest expected dependency are not in the traditional WIN-mandatory category. Thus, in this report, the "most dependent" subgroups do not, in fact, include these cases. Eligibility in the three programs was broad. San Diego worked with all mandatory applicants, Baltimore enrolled applicants and newly-mandatory recipients, and Virginia served the entire mandatory caseload. None, however, worked with AFDC recipients who had young children. Moreover, dependency in this study was measured by dollars of welfare received over a relatively short period: from one to at most three years following program enrollment. In addition, the data on which this analysis is based come from relatively low-cost programs that did not provide, for the most part, intensive services. Most importantly, no subgroups were singled out for special targeting attention.

Caution is urged in considering possible targeting options for this

WIN-mandatory population. Too narrowly defined targeting may destroy the value of certain services. Working with only a small subgroup may reduce overall effects on a caseload, even if subgroup effects are larger than average. A closely related question is "tracking" versus "mainstreaming," an issue widely discussed in education. A tracking agenda puts high- and low-achieving students into separate classes. Mainstreaming puts the two together, with the idea that the brighter students can assist the others. An open question in welfare employment programs is whether loosely structured, low-cost services, such as job search workshops, can be effective if women with no prior work experience do not have the opportunity to learn from others who have held jobs. Prior job-holders, who often find new jobs quickly, may, in addition, provide the necessary boost for other participants to keep trying. "Tracking," or separating out inexperienced workers, may also create staff problems if generally poor success rates demoralize staff instructors.

## C. The Demonstration of State Work/Welfare Initiatives

MDRC's Demonstration of State Work/Welfare Initiatives was launched in 1982 to test the effectiveness of state employment programs for people applying for or receiving AFDC. For the most part, states were using their new authority to experiment with WIN programs authorized by the Omnibus Budget Reconciliation Act (OBRA) of 1981. The MDRC study includes programs in 11 states, eight of which used random assignment to form experimental and control groups for full-scale impact and benefit-cost studies. Most programs have the goal of increasing employment and reducing the dependency of the welfare population by preparing recipients for work. Thus, most

able-bodied recipients had to participate in job search and/or unpaid work experience or other activities as a condition of welfare receipt.

The research was designed to assess three areas: the feasibility of implementing a mandatory participation and/or work requirement; the program's impacts on employment, earnings and welfare receipt; and the cost-effectiveness of the different approaches. Findings from this MDRC study are being released as the results for each state's program become available. The programs in this study are examined in more detail in Chapter 2.

In the three areas studied, the evaluations generally found that employment and earnings improved, and, in two areas, there were welfare savings. Also, the results for two of the programs (in San Diego and Virginia) indicated the initial investment of funds in the programs would result in government budget savings within a five-year time-frame or less.

The subgroup impacts in these evaluations have suggested the possibility of finding better methods to serve groups within the diverse welfare population. For example, employment increases have generally been larger for clients without a recent work history than for those who have worked during the year prior to program enrollment. These findings are buttressed by research conducted by MDRC in prior WIN programs and findings from the National Supported Work Demonstration.[11] This study is able to examine a wider variety of subgroups than were analyzed in the final reports, and uses longer-term data with a methodology more suited to the questions of performance measures than was possible in the previous evaluations.

## CHAPTER 2

## CHARACTERISTICS OF PROGRAMS AND PARTICIPANTS

This chapter discusses the similarities and differences between the three state programs examined in this paper: the San Diego, Baltimore and Virginia programs. The chapter then describes the characteristics of the research samples as well as some of the normal behavioral differences among welfare population subgroups in the absence of special services.

### A. The Program Models

No single program model was tested in MDRC's Work/Welfare study. Rather, the participating states implemented their own initiatives, using different strategies. Characteristics of the local WIN-mandatory populations often differed as well.

The evaluations, on the other hand, are similar in methodology: each study used an experimental design whereby program enrollees were randomly assigned to one or more experimental groups or to control groups. Experimental group members were subject to mandatory participation requirements (e.g., they were required to take part in program services), while the control groups were barred from the special programs, although in some areas they could receive the minimal WIN services that were offered. Data were collected on participation measures, outcomes in employment and welfare receipt, and direct program operating costs. To estimate program impacts, the employment and welfare behavior of the experimental and control groups were compared over several quarters of follow-up. Because

randomization had produced experimental and control groups with similar demographic characteristics and backgrounds in prior employment and welfare dependency, any statistically significant differences in behavior could be safely attributed to the programs' treatments.

In these studies, the term "applicant" identifies a person applying for AFDC at the time of entry into the research sample, whether or not that person's welfare grant was subsequently approved. That label remains, even when the person becomes a recipient. The term "recipient" refers to a sample member who was already receiving welfare at the date of sample entry. These two subgroups are important and are analyzed separately throughout much of this study. Other subgroup divisions are based on prior demographic and background characteristics.

Table 2.1 shows the key characteristics of the programs involved in this analysis. The published state reports contain more detail about both the programs and the evaluation results.[1] Briefly, job search and work experience -- along with education and training in Baltimore, and, to a lesser extent, in Virginia -- were the major program services, but states differed in the mix and intensity of these services, their sequencing, and the populations that received them. Programs were all mandatory, but differed in the extent to which participation was enforced.

San Diego worked with all WIN-mandatory welfare applicants but did not enroll recipients. Experimentals went through a two-stage fixed sequence of group job search followed by a 13-week work obligation, if they had not found unsubsidized jobs in the first phase.[2] San Diego's decision to focus on applicants rather than recipients represents one targeting option available to program operators.

TABLE 2.1

KEY CHARACTERISTICS OF STATE WORK/WELFARE INITIATIVES

| Characteristic | San Diego, California[a] | Baltimore, Maryland[b] | Virginia[a] |
|---|---|---|---|
| Eligible Group | | | |
| Applicants | Yes | Yes | Yes |
| Newly Mandatory Recipients | No | Yes | Yes |
| Currently Mandatory Recipients | No | No | Yes |
| Enrollment Limit | None | 1000/year | None |
| Program Model | Job search workshop followed by 13 weeks of CWEP in public and private nonprofit agencies | Multi-component, including job search, education, training, on-the job training and 13 weeks of work experience. | Job search followed by 13 weeks of CWEP, education or training |
| Sequence | Fixed: job search then work experience | Discretionary | Job search first |
| Client Choice of Components | No | Yes | Yes |
| Components | | | |
| Job Search | Mandatory | Mandatory when judged appropriate | Mandatory as first component |
| Independent | No | Yes | Yes |
| Group | Yes | Yes | Yes |
| Work Experience | Mandatory if no job found through job search | Mandatory when judged appropriate | Mandatory when judged appropriate |
| Education and Training | None | In-house and by referral | By referral |
| Study Area[c] | County-wide | 10 out of the 18 Income Maintenance Centers | 11 of 124 agencies (4 urban, 7 rural) |
| Research Method | Random assignment to either of two experimental groups. Controls get WIN services. | Random assignment. Controls get WIN services. | Random assignment to either of two experimental groups. Controls get no special services. |
| Sample Enrollment Period | October 1982 - August 1983 | November 1982 - December 1983 | August 1983 - September 1984 |

NOTES: [a]In San Diego and Virginia there are two different experimental treatments.

[b]In Maryland, a full evaluation was conducted in the indicated area and a process study was done in another area as well.

[c]In addition to the study areas, Virginia implemented the program statewide.

Baltimore, on the other hand, enrolled both WIN-mandatory applicants and recipients, but only recipients who had just become mandatory, usually because their youngest child had turned six years of age. In order to ensure adequate funding on an individual basis for a somewhat broader array of services, the Baltimore program restricted enrollment to only 1,000 registrants a year. The program provided a mix of components (including job search, unpaid work experience, education and training), and staff made service assignments according to enrollees' needs and preferences, depending on their assessments and the availability of open slots.

Virginia enrolled a sample representative of its entire existing WIN-mandatory caseload. The state stipulated that counties require job search of all enrollees but authorized, as a county option, short-term work experience, education and training as follow-up activities. Education and training were not provided by the program; rather, participants were referred to JTPA and community schools with independent funding, open to all who qualified. Consequently, control group members participated in education and training with a frequency equal to experimentals.

The treatments were relatively inexpensive, but did vary in average cost per experimental. For example, the cost of the San Diego program was about two-thirds that of the Baltimore program, which spent, on average, $1,050 per experimental. San Diego spent more on ensuring compliance with its participation requirement (which entailed monitoring, registrant follow-up and limited sanctioning), while Baltimore offered more expensive services, such as education and training, and provided client stipends.

The different funding levels and philosophies determined how mandatory -- as measured by participation and sanctioning rates -- each program was.[3]

In San Diego, less than one in 10 experimentals was not reached by the program: that is, few people were still on welfare, not employed, and had not participated in the program after nine months following program entry. This high San Diego coverage indicates that a short-term participation requirement was, in fact, realized by that program. In contrast, a larger proportion of registrants -- almost one-quarter -- were not involved in formal activities in the Baltimore program. Although this may be partly due to the mix of recipients with applicants, it also reflects the Baltimore staff's greater flexibility in deferring registrants. In Virginia, most experimentals were reached by the program (nearly 90 percent), but the minimum requirement -- a loosely structured form of independent job search -- was relatively easy for both the program and the clients to fulfill.

Local economic conditions, staff experience and attitudes also differed. Statutory grant maximums, based on state standards of need, also varied widely, making cross-program comparisons problematic. Low benefit levels increased the attractiveness of low-wage jobs in some areas, and also increased the likelihood of a case closure when employment was obtained. In San Diego, welfare recipients had a good market in which to look for jobs, but in rural areas of Virginia, the prospects for employment were limited. And, in the administrative reorganization permitted by OBRA, social service staffs in some states -- who had recently assumed new responsibility for employment functions -- had to go through a learning process. However, staffs in San Diego and Baltimore had substantial prior experience in operating employment and/or work programs, which contributed to their programs' smooth administration.

B. Sample Characteristics

Table 2.2 shows the size of the samples randomly assigned[4] in the San Diego, Baltimore and Virginia programs, while Table 2.3 describes sample characteristics, broken down by the subgroups analyzed in this paper.[5] Differences in the program models, targeting philosophies, and the environments in which the programs operated created variation. As noted earlier, each program served the WIN mandatory caseload (which excludes women with children less than six years of age), or portions of that caseload. The San Diego program served only applicants, while the Baltimore and Virginia samples had a fairly even mix of applicants and recipients, although the type of recipient differed.

The Baltimore and Virginia samples were similar in many respects: over half had neither a high school diploma nor a GED; more than half had been receiving AFDC for more than two years; and, on average, only about 40 percent had held a job in the year prior to random assignment. The San Diego sample was less disadvantaged. More than half were high school graduates; less than 30 percent had been on welfare for more than two years; and one-half had held a job in the year before this welfare application. Ethnic composition also differed. In Baltimore and Virginia, between 60 to 70 percent of the samples were black; in San Diego only 20 percent of sample members were black, although Hispanics made up 18 percent of the sample.

Comparisons of applicants and recipients in Baltimore and Virginia reveal large differences in prior earnings and prior welfare receipt -- in fact, applicants in all three states were remarkably similar as were

TABLE 2.2

DISTRIBUTION OF THE RESEARCH SAMPLE,
BY PERIOD OF RANDOM ASSIGNMENT AND WELFARE STATUS

| Program and Period of Random Assignment | Total | Applicants | Recipients |
|---|---|---|---|
| San Diego[a] | | | |
| October-December 1982 | 555 | 555 | n/a |
| January-March 1983 | 882 | 882 | n/a |
| April-June 1983 | 505 | 505 | n/a |
| July-August 1983 | 459 | 459 | n/a |
| Total | 2381 | 2381 | n/a |
| Baltimore | | | |
| November-December 1982 | 323 | 166 | 157 |
| January-March 1983 | 709 | 332 | 377 |
| April-June 1983 | 444 | 199 | 245 |
| July-September 1983 | 650 | 340 | 310 |
| October-December 1983 | 631 | 343 | 288 |
| Total | 2757 | 1380 | 1377 |
| Virginia | | | |
| August-September 1983 | 372 | 66 | 306 |
| October-December 1983 | 1105 | 337 | 768 |
| January-March 1984 | 797 | 326 | 471 |
| April-June 1984 | 507 | 276 | 231 |
| July-September 1984 | 369 | 264 | 105 |
| Total | 3150 | 1269 | 1881 |

SOURCE: MDRC tabulations from Client Information Sheets.

NOTES: N/A indicates not applicable because there were only applicants in San Diego.

[a]There were three research groups in San Diego: Controls, Job Search/Work Experience Experimentals, and Job Search Only Experimentals. Only the first two groups were used for the analysis of subgroup impacts in this report, and the sample size shown here refers to those two. However, percent of sample calculations for subgroups, shown in later tables, are based on the full sample (3238 individuals).

TABLE 2.3

SELECTED CHARACTERISTICS OF AFDC
APPLICANTS AND RECIPIENTS AT TIME OF RANDOM ASSIGNMENT,
BY PROGRAM AND WELFARE STATUS

| | San Diego | Baltimore | | | Virginia | | |
|---|---|---|---|---|---|---|---|
| Subgroup | Applicants | Applicants | Recipients | Total | Applicants | Recipients | Total |
| Prior Earnings (%) | | | | | | | |
| $3000 or More | 28.8 | 31.9 | 6.7 | 19.3 | 29.2 | 3.3 | 13.7 |
| $1-2999 | 22.9 | 29.3 | 20.6 | 24.9 | 28.4 | 19.7 | 23.3 |
| None | 48.4 | 38.8 | 72.8 | 55.8 | 42.3 | 77.0 | 63.0 |
| Had Own AFDC Case (%) | | | | | | | |
| Never | 33.4 | 22.7 | 5.2 | 14.0 | 26.2 | 2.5 | 12.1 |
| Two Years or Less | 38.7 | 41.8 | 21.1 | 31.4 | 31.7 | 25.7 | 28.1 |
| More Than Two Years | 27.9 | 35.5 | 73.8 | 54.6 | 42.2 | 71.8 | 59.8 |
| High School Diploma (%) | | | | | | | |
| Yes | 61.5 | 44.9 | 42.1 | 43.5 | 50.8 | 38.8 | 43.6 |
| No | 38.5 | 55.1 | 57.9 | 56.5 | 49.2 | 61.2 | 56.4 |
| Child 12 or Under (%) | | | | | | | |
| No | 22.6 | 27.5 | 13.4 | 20.5 | 22.9 | 23.7 | 23.4 |
| Yes | 77.4 | 72.5 | 86.6 | 79.5 | 77.1 | 76.3 | 76.6 |
| Number of Own Children (%) | | | | | | | |
| One | 49.7 | 50.4 | 43.1 | 46.8 | 49.6 | 42.0 | 45.0 |
| More Than One | 50.3 | 49.6 | 56.9 | 53.2 | 50.4 | 58.0 | 55.0 |
| Currently Married (%) | | | | | | | |
| Yes | 46.6 | 50.4 | 34.3 | 42.3 | 49.3 | 38.3 | 42.7 |
| No | 53.4 | 49.6 | 65.7 | 57.7 | 50.7 | 61.7 | 57.3 |
| Ever Married (%) | | | | | | | |
| Yes | 84.1 | 69.9 | 49.1 | 59.5 | 74.2 | 65.3 | 68.8 |
| No | 15.9 | 30.1 | 50.9 | 40.5 | 25.8 | 34.7 | 31.2 |
| Age (%) | | | | | | | |
| 30 or Over | 65.6 | 65.4 | 42.7 | 54.0 | 64.0 | 65.9 | 65.1 |
| Under 30 | 34.4 | 34.6 | 57.3 | 46.0 | 36.0 | 34.1 | 34.9 |
| Ethnicity (%) | | | | | | | |
| White | 61.5 | 35.1 | 25.9 | 30.5 | 43.7 | 28.8 | 34.8 |
| Black | 20.7 | 64.6 | 73.9 | 69.2 | 54.9 | 70.3 | 64.1 |
| Hispanic | 17.8 | 0.4 | 0.3 | 0.3 | 1.4 | 1.0 | 1.1 |
| Recent UI Benefits (%) | | | | | | | |
| Some | 14.1 | n/a | n/a | n/a | 3.9 | 0.4 | 1.8 |
| None | 85.9 | n/a | n/a | n/a | 96.1 | 99.6 | 98.2 |
| Labor Market (%) | | | | | | | |
| Urban | n/a | n/a | n/a | n/a | 78.7 | 78.8 | 78.8 |
| Rural | n/a | n/a | n/a | n/a | 21.3 | 21.2 | 21.2 |
| Sample Size | 3238[a] | 1380 | 1377 | 2757 | 1289 | 1881 | 3150 |

SOURCE: MDRC calculations from MDRC Client Information Sheets.

NOTES: Distributions may not add to exactly 100.0 percent due to rounding.

[a]As explained in Table 2.2, the percent of sample calculations are based on Job Search/Work Experience Experimentals, Job Search Only Experimentals and Controls, for a total of 3238. Elsewhere in this report, the sample size is 2381, since Job Search Only Experimentals were not included.

Tests of statistical significance were not calculated.

recipients in the two states that served them. Among applicants, somewhat less than one-third were first-time applicants; another one-third fell into the top earnings category ($3,000 or more for the pre-program year). In both Baltimore and Virginia, three-quarters of the recipients had been on welfare for more than two years.

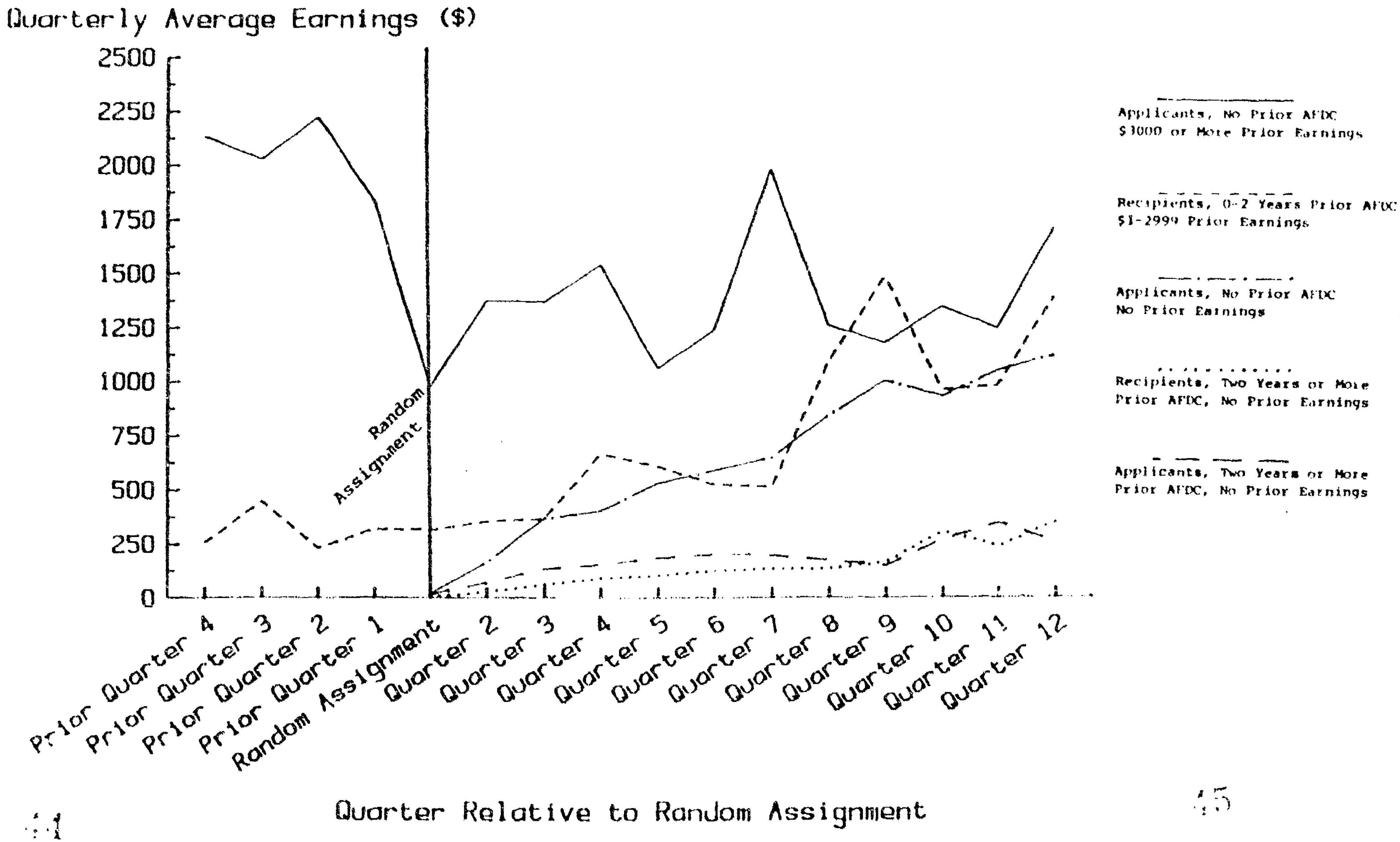
C. Earnings and Welfare Receipt: The Normal Range

A wide range of earnings and welfare behavior of WIN-mandatory clients in the absence of program intervention can be captured by simple objective measures obtained at the time of random assignment. Figures 2.1 and 2.2 plot the earnings and welfare receipt of the early Baltimore control sample by selected subgroups defined by applicant/recipient status, prior employment and prior welfare receipt.

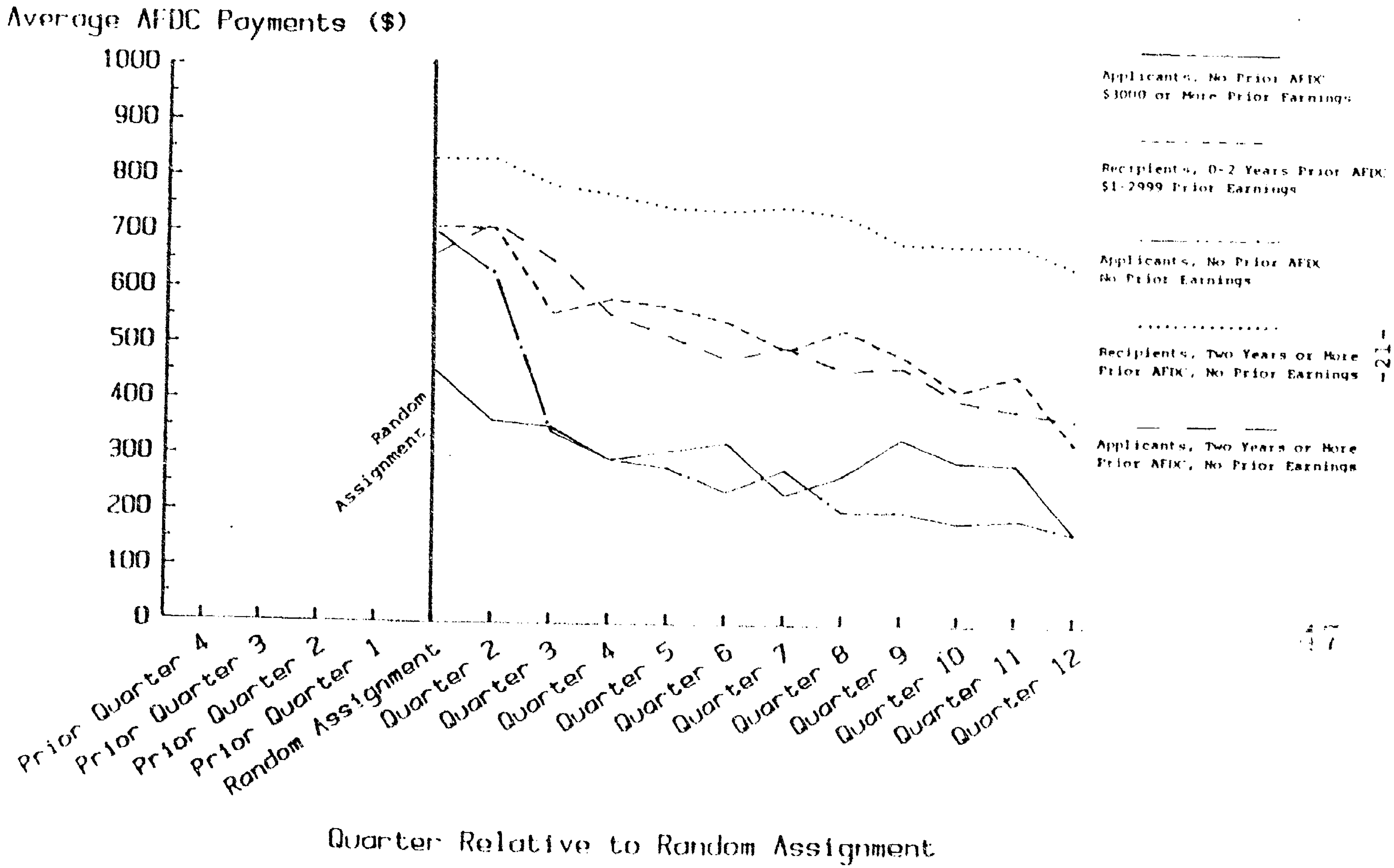
The subgroup differences in the Baltimore control sample were large. Quarterly average earnings for control group applicants without a prior welfare history and with $3,000 or more in earnings in the year prior to AFDC application consistently fell into the $1,200 to $1,800 per-quarter range after the first year of follow-up (counting zero earnings for persons not employed). During the same period, subgroups with no recent employment history and a pattern of AFDC receipt for more than two years barely averaged earnings of from $200 to $400 per quarter.

Welfare payments to control groups members also differed, depending on prior earnings and extent of previous welfare dependency. After three years, long-term recipients without pre-program earnings were receiving from three to four times the quarterly benefit payments of first-time applicants. Put another way, recipients for more than two years who had no

# FIGURE 2.1
# BALTIMORE CONTROLS, EARLY COHORT: QUARTERLY AVERAGE EARNINGS, BY SUBGROUP



# FIGURE 2.2
## BALTIMORE CONTROLS, EARLY COHORT: QUARTERLY AVERAGE AFDC PAYMENTS, BY SUBGROUP



earnings in the pre-program year were only one-third of the sample, but consumed nearly half of the total AFDC expenditures at the three-year mark.[6] A further breakdown of recipients by whether or not they had high school diplomas revealed that dropouts who had not obtained even a GED were 18.0 percent of the sample but received 28 percent of the AFDC dollars. In contrast, applicants -- about one-half of the sample -- were receiving less than one-third of all welfare payments in Baltimore.

These subgroups exhibit the full range of behavior connected with three major characteristics -- applicant/recipient status, prior earnings and prior AFDC receipt. The analyses to date suggest that these characteristics are some of the best predictors of future earnings and welfare receipt for program eligibles, and impact differences may well be associated with these measures. Subgroups defined by these three dimensions will be the subgroups with priority in this investigation.

## CHAPTER 3

## METHODOLOGY

This chapter reviews the principal elements of the experimental research design and the methodology used in this study. This chapter is meant as a general guide, although some of the discussion is of a more technical nature.

### A. Experimental Design

Any valid analysis of program impacts is based on a fundamental comparison between the observed outcomes of a program and what would have occurred without it. As explained in Chapter 1, program outcomes are relatively easy to observe. But the calculation of change -- or program impact -- requires estimates of outcomes in the absence of the program.

A classical experimental design is the preferred way of obtaining the standard for comparison. In such designs, clients are assigned on a random basis to either program services, the experimental group, or to a control group, which receives only the services available without the program. The average outcomes of experimentals, minus the average outcomes of controls, provide the program impact estimates, which show the program achievements over and above the normal job-finding and welfare patterns of the eligible population.

To maintain the integrity of the research design, no changes were made in the research group designations after random assignment. "Experimentals" remained experimentals and "controls" remained controls. In the calcula-

tion of outcomes, experimentals who did not, for some reason, participate in the programs were still counted as part of the experimental group. Their actions could influence program impacts, which are expressed on a _per experimental_ rather than on a _per participant_ basis. Nonparticipants, for instance, could take jobs or leave welfare on their own or be influenced by the program's participation requirement (since they could be sanctioned if they refused to participate).

The definition of subgroups follows this same labeling pattern. Subgroups are defined by _pre-existing_ characteristics at enrollment, not by any subsequent behavior or activity.

## B. Data Sources

Earnings and welfare data were assembled from administrative records. The use of such records offers several advantages. First, administrative records can be much less expensive than survey data, in part because registrants do not have to be re-contacted during the follow-up. Records may also be more accurate than survey data because they do not depend on client recall of dollar amounts of earnings or welfare payments. Different rates of response by the experimental versus the control group -- often a source of bias in survey data -- are also not expected with records data.

Administrative records are, however, limited in their comprehensiveness and coverage. For example, quarterly earnings information can be obtained from the Unemployment Insurance (UI) system, but data on wages and hours worked are not available. Moreover, the information can only be obtained with a lag, and some delinquency in filing earnings reports on the part of employers is common in wage-reporting states. Another drawback is

that state UI systems do not normally record the earnings of people who commute to work across state lines. Given random assignment, however, none of these factors should affect experimental and control group outcomes differently.

In addition, administrative records in this study contained no information on people other than the research sample members. They do not, for example, provide the earnings of other family members, whose income (both earned and unearned) will affect a household's welfare dependency and general well-being.

The completeness and accuracy of the records data collected in this study were examined by comparing a small sample of data from the analysis tapes to the original paper or microfilm documents in state or county offices. Earnings and welfare payments were well-matched. Further, a comparison of records and survey data from the Louisville WIN Laboratory and an earlier San Diego study suggests that the two sources yielded comparable information, although administrative records showed larger total welfare receipt than the self-reports in interviews.[1]

Records data were merged with demographic and program activity information to form a single program data base, with a new record compiled for each sample member. Each record contains the client's employment background and welfare history in addition to a series of outcome measures (quarterly UI earnings, monthly AFDC payments) running from the point of entry into the sample (i.e., the date of random assignment) through to the end of the follow-up. Program activities and dates are also included. The earlier a person entered the sample, the more follow-up data are available. No sample member has less than four quarters of earnings data and 12 months

of welfare data. The Baltimore program, with the longest -- and as yet unpublished follow-up data, has the most complete information -- an additional year of earnings and welfare beyond the results reported in that program's final report.

The major data sources for all the programs analysed are summarized below:[2]

- Client Information Sheets, one-page questionnaires filled out by client and staff as part of the random assignment process, provide information on the demographic characteristics of sample members. All principal subgroups, with the exception of the subgroups identified by prior earnings, were defined using this information.

- State Unemployment Insurance (UI) Earnings Records provide quarterly employment and earnings data reported by employers for each calendar quarter: e.g., January, February and March; April, May and June.

- AFDC records supply information on monthly AFDC (i.e., welfare) grants. Monthly AFDC data are grouped by three-month periods, where the first month of the first quarter of follow-up is the month of enrollment.

- Unemployment Insurance Benefit Records supply information on monthly UI benefit payments.

- Program Activity records provide information on program services, participation and deregistration.

Since random assignment can occur in the first, second, or third month of a calendar quarter, the first quarter of UI earnings can contain pre-program earnings for some sample members. The first quarter of earnings is therefore not considered a clean follow-up quarter in the impact analysis and is omitted from cumulative estimates of program impact.

C. Choice of Follow-up Period

MDRC's research to date has shown certain patterns of outcomes for

experimentals and controls over time. Typically, the outcomes for experimentals and controls were similar in the quarter of random assignment but began to differ in quarter 2. (However, many experimentals did not join activities for as long as six months after enrollment.) The experimental-control differences grew slowly, with the difference often peaking at the one-year point or beyond.

This paper divides follow-up into an immediate post-random assignment period (quarters 1 through 3) and a longer-term follow-up period (quarters 4 and following). Quarters were averaged -- which helps to eliminate some of the transitory quarter-to-quarter variation in earnings. Earnings, as well as employment, welfare incidence and AFDC payments, are expressed as quarterly averages per person. Averages for tha immediate and longer-term outcomes were calculated separately. It should be emphasized that the longer-term average contains more quarters of data for persons who entered the research early. This averaging procedure has the disadvantage that it does not explicitly estimate quarter-by-quarter time trends in impacts.

The longer-term follow-up period was selected as the focus of this subgroup analysis because it best represents both post-program outcomes and impacts. Subgroup differences appearing in the later quarters are the best indicators of long-run effects and are therefore likely to be more indicative of the total impact differences among subgroups. The training activities and education programs in Baltimore, which run in duration for as much as one year, require a long follow-up period, with an emphasis on later periods. Unfortunately, the follow-up in Virginia was only fou quarters for the substantial portion of the sample.

Statistical tests were conducted and are reported for differences

between experimentals and controls within subgroups. While the differences between impacts for pairs of subgroups were also tested, they were not as frequently statistically significant. The results of such tests are omitted from the tables but are occasionally mentioned as appropriate.

## D. The Subgroup Impact Regression Model

A simple difference between average outcomes for experimental and control groups is sufficient to express reliable impacts in a carefully implemented experimental design. The use of linear regression may, however, lend extra precision to the estimates and correct for minor differences in pre-program characteristics between experimentals and controls. For this reason, the estimates reported in this paper are regression-adjusted.

In addition, regression techniques have been used to produce two sets of subgroup impacts. The first set takes the point of view of the program administrator who asks: "Can I improve efficiency by targeting services to registrants with a single subgroup characteristic?" For example, it may be useful to find out if sample members with a high school diploma have different impacts than those without diplomas, ignoring differences in any other demographic characteristics. These impact estimates are unconditional estimates, and this type of estimate is the focus of Chapter 4. Such subgroup estimates do not take into account impact differences associated with other demographic and background characteristics. For example, women without a high school diploma generally have a weaker work record, but unconditional estimates do not explain what part of the diploma effect is due to the work history characteristic. Regression, in this case, serves only the purpose of increasing precision and adjusting for minor pre-existing experimental-

control differences.

Two or more characteristics can be included in unconditional estimation as _interactions_, and these are often useful to program operators. To continue the example above, the sample may be split four ways: persons with and without diploma, further divided by employed/not employed in the recent pre-program period. Impacts calculated for each of these four subgroups may answer the question as to whether it is worthwhile to target services to a narrow subgroup defined by diploma and prior employment status. This approach provides information about targeting on the basis of two subgroup characteristics, without controlling for other factors.

Regression analysis can lead to another set of estimates -- _conditional_ estimates -- that may reveal the associations of underlying factors. Conditional estimates hold all subgroup characteristics constant except the one in question. That is, any conditional impact difference associated with a high school diploma would indicate the importance of the schooling credential itself, eliminating effects due to prior employment record and other characteristics. If conditioning on prior employment status nullified the diploma effect, then the prior-employment difference across diploma subgroups may be considered the "real" reason for the diploma impact.[3]

Both unconditional and conditional estimates are important, depending on the questions asked. Unconditional estimates are presented and discussed in the next chapter because they address questions of targeting with limited information. Conditional estimates, however, are required for the testing of performance measures in Chapter 5. They will be discussed in Chapter 4 only insofar as they raise issues regarding the conclusions of

the main analysis.

## V. Testing Performance Indicators

A handful of prior studies have attempted to test the correlation between various measures of performance and net program impact. These studies did not have experimental comparison data, but their techniques are similar to the ones used in this study of performance measures.

The basic approach is as follows:

1. Obtain an estimate of net program impact for each individual in the treatment group;

2. Create a measure of program performance -- e.g., did the sample member enter employment, what were his/her wages?

3. Compute correlation coefficients between the net impact and the performance measures, with measures with the greatest correlation being identified as the "best" performance indicators;

4. As a supplemental analysis, determine whether two indicators work better than one. Compute a regression of net impact on two performance indicators and report the coefficients and their statistical significance. In this way, it may be possible to determine that one indicator has more power than another or is a useful supplement.

This procedure has remained approximately the same since studies correlated performance measures with the impacts of certain pre-CETA employment programs.[4]

The difficult part of this process is the first step: the estimation of a net impact for each individual.[5] Prior studies estimated individual-level impacts without experimental data, and thus have had to depend on impact estimates from participant/nonparticipant comparisons adjusted by regression for various demographic and participation variables, such as type of treatment and length of stay. Thus, while these studies have used

essentially the same procedure to estimate individual impacts, the estimates they have generated may be biased insofar as the regression models used were not able to control for all observable and unobservable differences between the participant and nonparticipant groups.

## CHAPTER 4

## SUBGROUP DIFFERENCES IN IMPACTS

This chapter summarizes the findings of an analysis of program impact differences for subgroups of the WIN-mandatory AFDC caseload in San Diego, Baltimore and Virginia. Using the data and statistical methods described in the last two chapters, the analysis develops estimates of employment and welfare impact differences and then explores some implications of those differences. Subgroup differences in program costs are also briefly considered. Additional results on the benefit-cost implications of the subgroup differences are available from MDRC.

The thrust of these findings is that when people were defined in terms of their prior work and welfare history, the least dependent WIN-mandatory applicants and recipients generally experienced below-average program impacts and often the smallest impacts overall. These findings suggest that a policy of targeting programs only to those in the WIN-mandatory caseload who are most "job-ready" would not be efficient. Impacts were much larger for subgroups who were less job-ready. There is some evidence, however, that focusing narrowly on only the groups who normally receive the most welfare may not be desirable -- at least with the relatively short-term, less intensive services offered by these three programs. As the chapter will delineate, these broad conclusions are drawn from a complex series of results and are subject to a number of important qualifications.

A. Composite Estimates

Before the results for each of the five experimental samples -- San Diego applicants, Baltimore applicants, Baltimore recipients, Virginia applicants, and Virginia recipients[1] -- are presented, the overall impact differences are examined collectively across the different programs and AFDC groups. (See Table 4.1.) The five sets of employment and welfare impacts have been combined into a single set of composite estimates as a summary device.[2] These estimates do not indicate the variation by program, but these differences are addressed later.

The impact estimates were calculated using data collected in the fourth quarter after random assignment and in all subsequent quarters through to the end of the follow-up periods. By the fourth quarter, most members of the experimental groups who were participating in the programs had already finished the activities or were no longer subject to the participation requirements. Thus, the impact estimates generally reflect the post-program experience of the samples, and are probably the best available indicators of the longer-term effects of these programs on the experimental groups.

1. Full Sample

As the composite estimates show, the average quarterly employment rate in this period was 38 percent for the experimental group compared to 34 percent for controls. The impact of 4 percentage points is statistically significant. Similarly, experimentals who worked earned an average of $1.679 per quarter which, taking into account individuals who did not work, totaled to $638 per quarter for all experimentals -- a figure that is $87 higher than the average earnings level of controls, for a 16 percent

TABLE 4.1

AFDC APPLICANTS AND RECIPIENTS: COMPOSITE IMPACTS ON EARNINGS AND WELFARE RECEIPT, BY MAJOR SUBGROUP

| Subgroup | Percent Employed Quarterly Quarters 4 - Last: Experimental | Control | Difference | Average Earnings Per Quarter Quarters 4 - Last ($): Experimental | Control | Difference |
|---|---|---|---|---|---|---|
| Full Sample | 38.0 | 33.9 | + 4.1*** | 638 | 550 | + 87*** |
| Prior Year Earnings | | | | | | |
| $3000 or More | 62.0 | 58.9 | + 3.1 | 1323 | 1235 | + 88 |
| $1-2999 | 48.3 | 45.0 | + 3.3* | 747 | 699 | + 48 |
| None | 26.0 | 21.1 | + 4.9*** | 397 | 291 | +107*** |
| Had Own AFDC Case | | | | | | |
| Never | 42.1 | 42.4 | - 0.3 | 828 | 811 | + 17 |
| Two Years or Less | 43.5 | 38.2 | + 5.2*** | 788 | 661 | +127*** |
| More Than Two Years | 33.8 | 29.0 | + 4.9*** | 494 | 406 | + 88*** |

| Subgroup | Percent Receiving AFDC Monthly Quarters 4 - Last: Experimental | Control | Difference | Average AFDC Payments Per Quarter Quarters 4 - Last ($): Experimental | Control | Difference |
|---|---|---|---|---|---|---|
| Full Sample | 47.6 | 48.9 | - 1.3 | 425 | 441 | - 16* |
| Prior Year Earnings | | | | | | |
| $3000 or More | 33.9 | 34.2 | - 0.3 | 301 | 309 | - 7 |
| $1-2999 | 44.2 | 45.0 | - 0.8 | 388 | 405 | - 18 |
| None | 52.5 | 54.7 | - 2.2* | 474 | 495 | - 21* |
| Had Own AFDC Case | | | | | | |
| Never | 30.9 | 33.1 | - 2.2 | 297 | 328 | - 31 |
| Two Years or Less | 40.6 | 41.6 | - 1.0 | 361 | 373 | - 12 |
| More Than Two Years | 57.4 | 58.8 | - 1.4 | 503 | 519 | - 17 |

SOURCE: MDRC calculations from the County of San Diego welfare records and Unemployment Insurance records from the EPP Information System; from the State of Maryland welfare and Unemployment Insurance records; and from the Commonwealth of Virginia Unemployment Insurance earnings records, welfare records from the Virginia Automated Client Information System, and Fairfax County AFDC case files.

NOTES: These composites are weighted averages of the regression-adjusted estimates for applicant and recipient categories in each state presented later in this chapter. The weights are the inverse of the estimated standard errors of the subgroup impact estimate. Dollar-denominated estimates include zero values for sample members not employed and for sample members not receiving welfare. There may be some discrepancies in calculating sums and differences due to rounding.

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.

improvement.

A clear subgroup pattern underlies these employment impacts for the full sample. Employment and earnings impacts for sample members with no earnings in the year prior to random assignment were larger than the overall average employment and earnings impacts, while the gains for the subgroups who did have prior earnings were smaller and usually not significant. The programs had raised the employment level of the less employable group by 5 percentage points compared to only 3 percentage points for the other two groups with prior earnings. This is the case not because members of the experimental group with better earnings records were less able to find employment than those with poorer records. In fact, they entered employment more frequently, as one would expect. Almost 50 percent of the experimentals with $1-2,999 in prior earnings, and 62 percent of those with even more, were employed during the follow-up period -- a much higher level than the 26 percent level achieved by experimentals with no previous earnings. But their employment was less of a gain over that of their control group counterparts.

Similarly, sample members who had been on welfare (i.e., had their own AFDC case) in the past showed significant gains in both employment and earnings, while those with less dependency experienced virtually no change. Controls with prior welfare had a lower employment rate on their own than those who had not been on welfare, but the more welfare-dependent experimentals made the greatest gains. Thus, using the simple measures of prior work and welfare experience to categorize individuals, the less employable and more dependent subgroups had the largest employment impacts.

A similar, although somewhat weaker, pattern can be seen in the

composite welfare impacts. Overall, the average proportion of individuals in the experimental group who received welfare each month was 1.3 percentage points lower than for controls; they also received $16 less in AFDC payments per quarter, with the latter impact statistically significant. The subgroup with no prior earnings had relatively high impacts.

2. Applicants and Recipients

Tables 4.2 and 4.3 break down these composite estimates for AFDC applicants and recipients. The applicant impacts are larger than those for recipients despite the fact that their participation rate in program services was somewhat lower. In fact, quarterly applicant employment impacts were about double those of recipients. A supplemental analysis -- in which applicant and recipient data were pooled and the impacts estimated separately for the two groups with demographic differences controlled -- indicated that the impact differences stemmed from the applicant/recipient distinction, and not from other factors.[3]

The separate subgroup results for applicants and recipients show a similar pattern to the overall estimates. Applicant impacts (shown in Table 4.2) indicate that the employment and earnings gains, as well as welfare savings, were lowest for first-time applicants -- those who reported never having had their own AFDC case. Some of these individuals may have received welfare on their mother's grant as a minor, but as a group they are clearly the least welfare-dependent. Impacts were also lowest for applicants with the best prior earnings records, with applicants in the two lower-earnings categories having significant impacts of similar magnitude. Similarly, employment and welfare impacts were largest (and statistically significant) for applicants who had been on welfare before,

TABLE 4.2

AFDC APPLICANTS: COMPOSITE IMPACTS
ON EARNINGS AND WELFARE RECEIPT, BY MAJOR SUBGROUP

| Subgroup | Percent Employed Quarterly Quarters 4 - Last | | | Average Earnings Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|
| | Experimental | Control | Difference | Experimental | Control | Difference |
| Full Sample | 44.7 | 39.4 | + 5.3*** | 851 | 721 | +130*** |
| Prior Year Earnings | | | | | | |
| $3000 or More | 62.7 | 59.8 | + 2.8 | 1345 | 1288 | + 57 |
| $1-2999 | 48.4 | 42.6 | + 5.8*** | 835 | 649 | +186*** |
| None | 29.8 | 23.5 | + 6.4*** | 519 | 379 | +141*** |
| Had Own AFDC Case | | | | | | |
| Never | 45.0 | 43.7 | + 1.3 | 947 | 910 | + 36 |
| Two Years or Less | 47.0 | 39.8 | + 7.1*** | 925 | 733 | +191*** |
| More Than Two Years | 41.8 | 35.5 | + 6.3*** | 694 | 551 | +143*** |

| Subgroup | Percent Receiving AFDC Monthly Quarters 4 - Last | | | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|
| | Experimental | Control | Difference | Experimental | Control | Difference |
| Full Sample | 35.0 | 37.3 | - 2.3** | 335 | 358 | - 23** |
| Prior Year Earnings | | | | | | |
| $3000 or More | 30.4 | 29.5 | + 0.8 | 276 | 271 | + 5 |
| $1-2999 | 35.3 | 39.0 | - 3.7* | 336 | 378 | - 42* |
| None | 37.8 | 41.4 | - 3.6** | 374 | 404 | - 30* |
| Had Own AFDC Case | | | | | | |
| Never | 25.4 | 26.7 | - 1.2 | 251 | 259 | - 8 |
| Two Years or Less | 33.6 | 36.7 | - 3.1* | 323 | 359 | - 36* |
| More Than Two Years | 44.2 | 46.5 | - 2.3 | 415 | 435 | - 20 |

SOURCE AND NOTES: See Table 4.1.

TABLE 4.3

AFDC RECIPIENTS: COMPOSITE IMPACTS
ON EARNINGS AND WELFARE RECEIPT, BY MAJOR SUBGROUP

| | Percent Employed Quarterly Quarters 4 - Last | | | Average Earnings Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|
| Subgroup | Experimental | Control | Difference | Experimental | Control | Difference |
| Full Sample | 29.3 | 26.6 | + 2.7** | 415 | 371 | + 43 |
| Prior Year Earnings | | | | | | |
| $3000 or More[a] | — | — | — | — | — | — |
| $1-2999 | 48.2 | 48.7 | - 0.5 | 644 | 758 | -113* |
| None | 22.2 | 18.7 | + 3.5** | 300 | 220 | + 80** |
| Had Own AFDC Case | | | | | | |
| Never[a] | — | — | — | — | — | — |
| Two Years or Less | 37.8 | 35.8 | + 2.1 | 608 | 566 | + 42 |
| More Than Two Years | 26.5 | 23.2 | + 3.4** | 349 | 301 | + 49 |

| | Percent Receiving AFDC Monthly Quarters 4 - Last | | | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|
| Subgroup | Experimental | Control | Difference | Experimental | Control | Difference |
| Full Sample | 66.2 | 66.0 | + 0.3 | 541 | 548 | - 7 |
| Prior Year Earnings | | | | | | |
| $3000 or More[a] | — | — | — | — | — | — |
| $1-2999 | 59.1 | 54.9 | + 4.1 | 462 | 444 | + 18 |
| None | 69.3 | 69.8 | - 0.5 | 572 | 583 | - 11 |
| Had Own AFDC Case | | | | | | |
| Never[a] | — | — | — | — | — | — |
| Two Years or Less | 53.8 | 50.8 | + 3.0 | 423 | 396 | + 27 |
| More Than Two Years | 70.8 | 71.2 | - 0.4 | 582 | 596 | - 14 |

SOURCE AND NOTES: See Table 4.1.

ADDITIONAL NOTE: [a]Dashes indicate that calculations represent less than 10 percent of the sample, and therefore are considered unreliable measures. Outcomes for the indicated samples are (reading across the table):

| | | | | |
|---|---|---|---|---|
| Prior Earnings = $3000 or More | | | | |
| Employment and Earnings | 59.3 | 55.6 | $1265 | $1092 |
| AFDC Receipt and Payments | 47.3 | 52.2 | $ 386 | $ 435 |
| No Prior AFDC History | | | | |
| Employment and Earnings | 31.5 | 37.4 | $ 482 | $ 521 |
| AFDC Receipt and Payments | 53.5 | 59.8 | $ 461 | $ 576 |

although the length of time on welfare does not seem to have mattered much.

Among recipients, very few people had high prior earnings or had never had their own AFDC case, making the corresponding subgroup impact estimates too imprecise to be reliable; results for these groups have thus been dropped from the table. The great majority of recipients fell into the "no prior earnings" and "more than two years of welfare" categories. Recipients without earnings in the previous year had statistically significant employment and earnings gains, although the gains were smaller than for applicants in the same subgroup. On the other hand, recipients with modest prior earnings had no gains at all -- in fact, they registered an earnings loss. On the welfare side, savings were small, although subgroup differences were in the same direction as the earnings gains.

The composite estimates thus provide dichotomous evidence on program effectiveness. For applicants, program services were generally effective for everyone except the most employable subgroups, for whom the services made very little difference. Interestingly, the impacts on the moderately and very dependent subgroups were about the same. However, program services were generally less effective for recipients, who typically include the most dependent individuals of all.

Characteristics other than prior earnings and welfare receipt -- such as education and marital status -- were less consistent across program samples. They sometimes appeared important, however, in determining the different subgroup impacts of different programs.

## B. Estimates for the Five Samples

In most cases, the patterns noted in the composite estimates hold up

across the five samples in the analysis, although there were some inconsistencies across program samples. There may be some important interactions between subgroup characteristics, program features and economic conditions that determine the local patterns. While it is too early in the research to analyze such interactions, the available evidence from these samples does suggest some promising directions for future research.

The following three tables present subgroup impact estimates for each of the five samples. Tables 4.4 and 4.5 show impacts for the same two sets of subgroups considered in the composite estimates, while Table 4.6 considers impacts on earnings and welfare payments associated with other subgroup characteristics. As with the composite results, the impact estimates start at the fourth quarter after random assignment and go through to the end of the observation period; estimates for the first three quarters, as well as for the follow-up period as a whole, were also calculated and are available from MDRC. The short-term results are generally consistent with the longer-term impacts, although their magnitude varied in some program settings.

1. San Diego

As Chapter 2 indicated, the welfare employment program in San Diego differed from the others in two important respects. First, the program served only welfare applicants. Second, all enrollees had the same short-term sequence of program activities -- job search followed by work experience. Moreover, participation rates were high for all subgroups.

The San Diego findings clearly indicate that the program had its greatest impacts on the less job-ready and more welfare dependent applicants. Those with the lowest prior earnings (zero dollars for the

TABLE 4.4

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EMPLOYMENT AND EARNINGS, BY PROGRAM, MAJOR SUBGROUP, AND WELFARE STATUS

| Subgroup, Welfare Status, and Program | Percent of Sample | Percent Employed Quarterly Quarters 4 - Last | | | Average Earnings Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| Welfare Status (%) | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 100.0 | 41.9 | 37.4 | + 4.5*** | 891 | 773 | + 118** |
| Baltimore | 50.1 | 46.5 | 42.2 | + 4.3** | 997 | 825 | + 172*** |
| Virginia | 40.3 | 46.5 | 38.9 | + 7.6*** | 682 | 576 | + 105* |
| Recipients | | | | | | | |
| Baltimore | 49.9 | 31.1 | 28.3 | + 2.8 | 472 | 436 | + 37 |
| Virginia | 59.7 | 27.4 | 24.9 | + 2.6 | 364 | 315 | + 49 |
| Prior Year Earnings | | | | | | | |
| $3,000 or More | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 28.8 | 60.7 | 58.7 | + 1.9 | 1444 | 1482 | - 39 |
| Baltimore | 31.9 | 65.0 | 62.5 | + 2.5 | 1453 | 1435 | + 18 |
| Virginia | 29.2 | 62.6 | 58.0 | + 4.7 | 1145 | 954 | + 192* |
| Recipients[a] | | | | | | | |
| Baltimore | 6.7 | — | — | — | — | — | — |
| Virginia | 3.3 | — | — | — | — | — | — |
| $1-2999 | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 22.9 | 42.9 | 41.7 | + 1.2 | 813 | 729 | + 84 |
| Baltimore | 29.3 | 51.5 | 45.0 | + 6.5* | 1068 | 729 | + 339*** |
| Virginia | 28.4 | 52.1 | 40.6 | +11.5** | 644 | 499 | + 146 |
| Recipients | | | | | | | |
| Baltimore | 20.6 | 46.9 | 51.0 | - 4.1 | 691 | 935 | - 245*** |
| Virginia | 19.7 | 49.6 | 46.4 | + 3.2 | 603 | 601 | + 3 |
| None | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 48.4 | 30.3 | 22.8 | + 7.5*** | 601 | 375 | + 225*** |
| Baltimore | 38.8 | 27.7 | 23.6 | + 4.1 | 569 | 398 | + 171* |
| Virginia | 42.3 | 31.6 | 24.4 | + 7.2* | 386 | 367 | + 19 |
| Recipients | | | | | | | |
| Baltimore | 72.8 | 23.9 | 19.1 | + 4.7** | 347 | 236 | + 112** |
| Virginia | 77.0 | 20.5 | 18.3 | + 2.2 | 259 | 207 | + 52 |

(continued)

TABLE 4.4 (continued)

| Subgroup, Welfare Status, and Program | Percent of Sample | Percent Employed Quarterly Quarters 4 - Last | | | Average Earnings Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| Had Own AFDC Case | | | | | | | |
| Never | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 33.4 | 44.1 | 41.8 | + 2.3 | 1018 | 981 | + 37 |
| Baltimore | 22.7 | 46.8 | 47.1 | - 0.4 | 1136 | 1015 | + 121 |
| Virginia | 26.2 | 44.5 | 42.9 | + 1.6 | 710 | 743 | - 33 |
| Recipients[a] | | | | | | | |
| Baltimore | 5.2 | — | — | — | — | — | — |
| Virginia | 2.5 | — | — | — | — | — | — |
| Two Years or Less | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 38.7 | 41.4 | 35.3 | + 6.1** | 898 | 732 | + 165** |
| Baltimore | 41.8 | 51.7 | 45.9 | + 5.8* | 1109 | 940 | + 169* |
| Virginia | 31.7 | 49.5 | 38.7 | +10.8** | 771 | 527 | + 245** |
| Recipients | | | | | | | |
| Baltimore | 21.1 | 44.5 | 38.4 | + 6.0 | 777 | 636 | + 141 |
| Virginia | 25.7 | 31.3 | 32.9 | - 1.6 | 471 | 509 | - 38 |
| More Than Two Years | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 27.9 | 40.0 | 35.2 | + 4.8 | 731 | 585 | + 146 |
| Baltimore | 35.5 | 40.6 | 35.0 | + 5.6* | 776 | 568 | + 208** |
| Virginia | 42.2 | 45.5 | 36.4 | + 9.1** | 596 | 508 | + 88 |
| Recipients | | | | | | | |
| Baltimore | 73.8 | 26.9 | 24.5 | + 2.4 | 371 | 364 | + 6 |
| Virginia | 71.8 | 26.2 | 21.8 | + 4.4** | 330 | 243 | + 87** |

SOURCE: See Table 4.1.

SAMPLE SIZE: Sample sizes are as follows: Applicants - San Diego = 2381, Baltimore = 1360; Virginia = 1269; and Recipients - Baltimore = 1377, Virginia = 1881.

NOTES: These data are regression-adjusted using ordinary least squares, controlling for pre-random assignment characteristics of sample members. Dollar-denominated estimates include zero values for sample members not employed and for sample members not receiving welfare. Regressions were run separately for applicants and recipients in each program. There may be some discrep[illegible]ies in calculating sums and differences due to rounding.

[a]Dashes indicate that calculations represent less than 10 percent of the sample, and therefore are considered unreliable measures. Outcomes for the indicated samples are (reading across the table):

| | | | | |
|---|---|---|---|---|
| Prior Earnings = $3000 or More | | | | |
| Recipients, Baltimore | 61.6 | 58.4 | $1175 | $1097 |
| Virginia | 57.1 | 51.4 | $1379 | $1085 |
| No Prior AFDC History | | | | |
| Recipients, Baltimore | 37.0 | 41.8 | $ 673 | $ 628 |
| Virginia | 23.3 | 31.2 | $ 234 | $ 381 |

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.

TABLE 4.5

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON
AFDC INCIDENCE AND PAYMENTS, BY PROGRAM, MAJOR SUBGROUP, AND WELFARE STATUS

| Subgroup, Welfare Status and Program | Percent of Sample | Percent Receiving AFDC Monthly Quarters 4 – Last | | | Average AFDC Payments Per Quarter Quarters 4 – Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| Welfare Status (%) | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 100.0 | 32.3 | 34.0 | – 1.7 | 436 | 469 | – 33 |
| Baltimore | 50.1 | 43.0 | 45.4 | – 2.4 | 366 | 380 | – 14 |
| Virginia | 40.3 | 29.8 | 32.8 | – 3.1 | 227 | 250 | – 23 |
| Recipients | | | | | | | |
| Baltimore | 49.9 | 70.2 | 70.2 | + 0.0 | 627 | 622 | + 5 |
| Virginia | 59.7 | 62.2 | 61.8 | + 0.4 | 468 | 485 | – 17 |
| Prior Year Earnings | | | | | | | |
| $3,000 or More | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 28.8 | 26.5 | 25.9 | + 0.6 | 326 | 323 | + 3 |
| Baltimore | 31.9 | 36.5 | 34.8 | + 1.7 | 295 | 288 | + 7 |
| Virginia | 29.2 | 28.6 | 28.3 | + 0.3 | 221 | 216 | + 5 |
| Recipients[a] | | | | | | | |
| Baltimore | 6.7 | — | — | — | — | — | — |
| Virginia | 3.3 | — | — | — | — | — | — |
| $1–2999 | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 22.9 | 30.7 | 33.7 | – 2.9 | 409 | 471 | – 63 |
| Baltimore | 29.3 | 42.2 | 48.0 | – 5.8 | 367 | 405 | – 38 |
| Virginia | 28.4 | 33.2 | 35.5 | – 2.3 | 254 | 286 | – 32 |
| Recipients | | | | | | | |
| Baltimore | 20.6 | 64.6 | 58.3 | + 6.3 | 542 | 499 | + 44 |
| Virginia | 19.7 | 53.6 | 51.6 | + 2.0 | 396 | 398 | – 3 |
| None | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 48.4 | 36.5 | 38.9 | – 2.4 | 514 | 554 | – 40 |
| Baltimore | 38.3 | 48.8 | 51.9 | – 3.2 | 424 | 437 | – 13 |
| Virginia | 42.3 | 28.2 | 34.1 | – 5.9* | 212 | 250 | – 38 |
| Recipients | | | | | | | |
| Baltimore | 72.8 | 73.1 | 74.3 | – 1.2 | 665 | 667 | – 3 |
| Virginia | 77.0 | 65.5 | 65.3 | + 0.2 | 495 | 514 | – 19 |

(continued)

TABLE 4.5 (continued)

| Subgroup, Welfare Status and Program | Percent of Sample | Percent Receiving AFDC Monthly Quarters 4 - Last | | | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| Had Own AFDC Case | | | | | | | |
| Never | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 33.4 | 22.7 | 23.0 | - 0.2 | 314 | 320 | - 5 |
| Baltimore | 22.7 | 33.3 | 35.8 | - 2.5 | 287 | 295 | - 9 |
| Virginia | 29.2 | 21.5 | 23.0 | - 1.5 | 163 | 173 | - 11 |
| Recipients[a] | | | | | | | |
| Baltimore | 5.2 | — | — | — | — | — | — |
| Virginia | 2.5 | — | — | — | — | — | — |
| Two Years or Less | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 38.7 | 33.1 | 36.9 | - 3.7 | 436 | 510 | - 74* |
| Baltimore | 41.8 | 40.2 | 40.9 | - 0.7 | 344 | 352 | - 8 |
| Virginia | 31.7 | 25.9 | 31.2 | - 5.2 | 202 | 238 | - 36 |
| Recipients | | | | | | | |
| Baltimore | 21.1 | 52.6 | 56.9 | - 4.3 | 451 | 475 | - 24 |
| Virginia | 25.7 | 54.5 | 45.0 | + 9.5** | 401 | 336 | + 65** |
| More Than Two Years | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 27.9 | 42.6 | 43.0 | - 0.4 | 580 | 589 | - 8 |
| Baltimore | 35.5 | 52.3 | 56.7 | - 4.4 | 443 | 468 | - 25 |
| Virginia | 42.2 | 37.8 | 40.2 | - 2.4 | 285 | 307 | - 22 |
| Recipients | | | | | | | |
| Baltimore | 73.8 | 75.7 | 74.2 | + 1.5 | 680 | 662 | + 18 |
| Virginia | 71.8 | 65.8 | 68.2 | - 2.4 | 498 | 539 | - 41** |

SOURCE: See Table 4.1.

SAMPLE SIZE: See Table 4.4.

NOTES: These data are regression-adjusted using ordinary least squares, controlling for pre-random assignment characteristics of sample members. Dollar-denominated estimates include zero values for sample members not employed and for sample members not receiving welfare. Regressions were run separately for applicants and recipients in each program. There may be some discrepancies in calculating sums and differences due to rounding.

[a]Dashes indicate that calculations represent less than 10 percent of the sample, and therefore are considered unreliable measures. Outcomes for the indicated samples are (reading across the table):

| | | | | |
|---|---|---|---|---|
| Prior Earnings = $3000 or More | | | | |
| Recipients, Baltimore | 54.6 | 59.6 | $477 | $503 |
| Virginia | 36.8 | 41.5 | $277 | $352 |
| No Prior AFDC History | | | | |
| Recipients, Baltimore | 63.8 | 66.1 | $601 | $659 |
| Virginia | 39.5 | 50.6 | $291 | $474 |

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.

TABLE 4.6

AFDC APPLICANTS AND RECIPIENTS: UNCONDITIONAL IMPACTS ON EARNINGS AND AFDC PAYMENTS, BY PROGRAM, MINOR SUBGROUP, AND WELFARE STATUS

| Subgroup, Welfare Status and Program | Percent of Sample | Average Earnings Per Quarter Quarters 4 - Last ($) | | | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| High School Diploma | | | | | | | |
| Yes | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 61.5 | 1068 | 923 | + 146** | 375 | 420 | - 44 |
| Baltimore | 44.9 | 1189 | 1106 | + 82 | 337 | 337 | - 0 |
| Virginia | 50.8 | 774 | 707 | + 68 | 215 | 206 | + 9 |
| Recipients | | | | | | | |
| Baltimore | 42.1 | 645 | 598 | + 46 | 557 | 546 | + 10 |
| Virginia | 38.8 | 501 | 434 | + 67 | 445 | 424 | + 21 |
| No | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 38.5 | 609 | 534 | + 74 | 532 | 547 | - 15 |
| Baltimore | 55.1 | 829 | 593 | + 236*** | 391 | 416 | - 25 |
| Virginia | 48.2 | 586 | 442 | + 144* | 239 | 296 | - 57** |
| Recipients | | | | | | | |
| Baltimore | 57.9 | 347 | 317 | + 30 | 679 | 677 | + 2 |
| Virginia | 61.2 | 277 | 240 | + 37 | 484 | 526 | - 42** |
| Child 12 or Under | | | | | | | |
| No | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 22.6 | 1001 | 678 | + 323*** | 293 | 341 | - 48 |
| Baltimore | 27.5 | 954 | 942 | + 12 | 256 | 244 | + 12 |
| Virginia | 22.9 | 665 | 533 | + 132 | 175 | 183 | - 9 |
| Recipients | | | | | | | |
| Baltimore | 13.4 | 271 | 242 | + 28 | 438 | 460 | - 22 |
| Virginia | 23.7 | 317 | 271 | + 46 | 356 | 370 | - 14 |
| Yes | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 77.4 | 858 | 803 | + 55 | 477 | 506 | - 29 |
| Baltimore | 72.5 | 1011 | 780 | + 232*** | 408 | 432 | - 24 |
| Virginia | 77.1 | 687 | 580 | + 97 | 242 | 270 | - 28 |
| Recipients | | | | | | | |
| Baltimore | 86.6 | 504 | 466 | + 38 | 657 | 647 | + 10 |
| Virginia | 76.3 | 378 | 328 | + 50 | 503 | 521 | - 18 |

(continued)

TABLE 4.5 (continued)

| Subgroup, Welfare Status and Program | Percent of Sample | Average Earnings Per Quarter Quarters 4 - Last ($) | | | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| Number of Own Children | | | | | | | |
| One | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 49.7 | 887 | 807 | + 80 | 346 | 355 | - 9 |
| Baltimore | 50.4 | 1033 | 766 | + 267*** | 299 | 327 | - 28 |
| Virginia | 49.6 | 652 | 495 | + 157** | 204 | 226 | - 23 |
| Recipients | | | | | | | |
| Baltimore | 43.1 | 528 | 493 | + 35 | 503 | 522 | - 19 |
| Virginia | 42.0 | 383 | 347 | + 36 | 378 | 390 | - 12 |
| More Than One | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 50.3 | 895 | 740 | + 155** | 525 | 580 | - 56* |
| Baltimore | 49.6 | 966 | 891 | + 75 | 434 | 434 | + 0 |
| Virginia | 50.4 | 710 | 659 | + 51 | 250 | 273 | - 23 |
| Recipients | | | | | | | |
| Baltimore | 56.9 | 430 | 392 | + 38 | 721 | 696 | + 24 |
| Virginia | 58.0 | 353 | 285 | + 67 | 534 | 554 | - 19 |
| Currently Married | | | | | | | |
| Yes | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 46.6 | 852 | 750 | + 102 | 469 | 450 | + 19 |
| Baltimore | 50.4 | 951 | 829 | + 122 | 362 | 374 | - 12 |
| Virginia | 49.3 | 678 | 504 | + 174** | 208 | 206 | + 1 |
| Recipients | | | | | | | |
| Baltimore | 34.3 | 436 | 422 | + 14 | 614 | 617 | - 3 |
| Virginia | 38.3 | 353 | 343 | + 10 | 491 | 471 | + 20 |
| No | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 53.4 | 925 | 794 | + 131* | 406 | 482 | - 76** |
| Baltimore | 49.6 | 1046 | 824 | + 222*** | 370 | 387 | - 16 |
| Virginia | 50.7 | 687 | 649 | + 37 | 245 | 293 | - 48* |
| Recipients | | | | | | | |
| Baltimore | 65.7 | 490 | 442 | + 48 | 634 | 624 | + 10 |
| Virginia | 61.7 | 371 | 299 | + 72 | 454 | 492 | - 38* |

(continued)

TABLE 4.6 (continued)

| Subgroup, Welfare Status and Program | Percent of Sample | Average Earnings Per Quarter Quarters 4 - Last ($) | | | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| Ever Married | | | | | | | |
| Yes | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 84.1 | 908 | 806 | + 102* | 421 | 445 | - 24 |
| Baltimore | 69.9 | 1003 | 821 | + 182** | 346 | 363 | - 17 |
| Virginia | 74.2 | 702 | 566 | + 136** | 205 | 212 | - 8 |
| Recipients | | | | | | | |
| Baltimore | 49.1 | 458 | 425 | + 33 | 621 | 624 | - 2 |
| Virginia | 65.3 | 369 | 308 | + 61 | 456 | 454 | + 2 |
| No | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 15.9 | 805 | 608 | + 196 | 511 | 591 | - 81 |
| Baltimore | 30.1 | 980 | 832 | + 148 | 413 | 421 | - 8 |
| Virginia | 25.8 | 625 | 610 | + 15 | 290 | 360 | - 70* |
| Recipients | | | | | | | |
| Baltimore | 50.9 | 487 | 446 | + 40 | 633 | 620 | + 13 |
| Virginia | 34.7 | 354 | 326 | + 28 | 491 | 542 | - 51* |
| Age | | | | | | | |
| 30 or Over | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 65.6 | 986 | 776 | + 163*** | 411 | 464 | - 54* |
| Baltimore | 65.4 | 1076 | 927 | + 150* | 349 | 347 | + 2 |
| Virginia | 64.0 | 714 | 580 | + 134** | 223 | 226 | - 3 |
| Recipients | | | | | | | |
| Baltimore | 42.7 | 398 | 394 | + 4 | 636 | 632 | + 3 |
| Virginia | 65.9 | 321 | 317 | + 4 | 476 | 486 | - 10 |
| Less Than 30 | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 34.4 | 747 | 770 | - 22 | 484 | 477 | + 7 |
| Baltimore | 34.6 | 846 | 633 | + 213** | 398 | 442 | - 44 |
| Virginia | 36.0 | 625 | 572 | + 53 | 235 | 295 | - 60* |
| Recipients | | | | | | | |
| Baltimore | 57.3 | 527 | 467 | + 61 | 621 | 614 | + 7 |
| Virginia | 34.1 | 448 | 316 | + 132** | 453 | 483 | - 30 |

(continued)

TABLE 4.8 (continued)

| Subgroup, Welfare Status and Program | Percent of Sample | Average Earnings Per Quarter Quarters 4 - Last ($) | | | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| Ethnicity | | | | | | | |
| White | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 61.5 | 949 | 821 | + 128* | 357 | 369 | - 12 |
| Baltimore | 35.1 | 923 | 760 | + 163 | 302 | 322 | - 20 |
| Virginia | 43.7 | 712 | 558 | + 154* | 166 | 189 | - 23 |
| Recipients | | | | | | | |
| Baltimore | 25.9 | 449 | 430 | + 19 | 578 | 577 | - 1 |
| Virginia | 28.8 | 378 | 338 | + 40 | 389 | 383 | + 6 |
| Black | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 20.7 | 895 | 589 | + 306*** | 532 | 678 | - 146*** |
| Baltimore | 64.6 | 1043 | 864 | + 180** | 400 | 411 | - 11 |
| Virginia | 54.9 | 657 | 581 | + 76 | 279 | 303 | - 24 |
| Recipients | | | | | | | |
| Baltimore | 73.9 | 484 | 439 | + 45 | 647 | 639 | + 8 |
| Virginia | 70.3 | 353 | 307 | + 46 | 503 | 530 | - 27 |
| Hispanic | | | | | | | |
| Applicants[a] | | | | | | | |
| San Diego | 17.8 | 693 | 843 | - 150 | 593 | 556 | + 38 |
| Baltimore | 0.4 | — | — | — | — | — | — |
| Virginia | 1.4 | — | — | — | — | — | — |
| Recipients[a] | | | | | | | |
| Baltimore | 0.3 | — | — | — | — | — | — |
| Virginia | 1.0 | — | — | — | — | — | — |
| Recent UI Benefits | | | | | | | |
| Some | | | | | | | |
| Applicants[a] | | | | | | | |
| San Diego | 14.1 | 1270 | 1304 | - 34 | 390 | 471 | - 80 |
| Baltimore | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Virginia | 3.9 | — | — | — | — | — | — |
| Recipients[a] | | | | | | | |
| Baltimore | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Virginia | 0.4 | — | — | — | — | — | — |
| None | | | | | | | |
| Applicants | | | | | | | |
| San Diego | 85.9 | 828 | 684 | + 144*** | 443 | 467 | - 24 |
| Baltimore | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Virginia | 96.1 | 666 | 565 | + 102* | 226 | 253 | - 27 |
| Recipients | | | | | | | |
| Baltimore | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Virginia | 99.6 | 363 | 311 | + 52 | 489 | 486 | - 16 |

(continued)

TABLE 4.6 (continued)

| Subgroup, Welfare Status and Program | Percent of Sample | Average Earnings Per Quarter Quarters 4 - Last ($) | | | Average AFDC Payments Per Quarter Quarters 4 - Last ($) | | |
|---|---|---|---|---|---|---|---|
| | | Experimental | Control | Difference | Experimental | Control | Difference |
| Labor Market | | | | | | | |
| Urban | | | | | | | |
| Applicants | | | | | | | |
| San Diego | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Baltimore | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Virginia | 78.7 | 705 | 603 | + 102* | 232 | 259 | - 27 |
| Recipients | | | | | | | |
| Baltimore | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Virginia | 78.8 | 401 | 331 | + 71* | 471 | 500 | - 29 |
| Rural | | | | | | | |
| Applicants | | | | | | | |
| San Diego | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Baltimore | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Virginia | 21.3 | 595 | 478 | + 117 | 206 | 216 | - 11 |
| Recipients | | | | | | | |
| Baltimore | n/a | n/a | n/a | n/a | n/a | n/a | n/a |
| Virginia | 21.2 | 223 | 252 | - 29 | 460 | 434 | + 26 |

SOURCE: See Table 4.1.

SAMPLE SIZE: See Table 4.4.

NOTES: These data are regression-adjusted using ordinary least squares, controlling for pre-random assignment characteristics of sample members. Dollar-denominated estimates include zero values for sample members not employed and for sample members not receiving welfare. Regressions were run seperately for applicants and recipients in each program. There may be some discrepancies in calculating sums and differences due to rounding.

N/A indicates not applicable because data was not available for these samples.

[a]Dashes indicate that calculations represent less than 10 percent of the sample, and therefore are considered unreliable measures. Outcomes for the indicated samples are in dollars (reading across the table):

| | | | | |
|---|---|---|---|---|
| Hispanic - | | | | |
| Applicants, Baltimore | -313 | 209 | 666 | 528 |
| Virginia | 719 | 876 | 48 | 70 |
| Recipients, Baltimore | -253 | 255 | 227 | 409 |
| Virginia | 742 | 233 | 310 | 276 |
| Some Recent UI Benefits - | | | | |
| Applicants, Virginia | 1058 | 867 | 245 | 169 |
| Recipients, Virginia | 618 | 1246 | 220 | 394 |

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.



BEST COPY AVAILABLE

year prior) had by far the largest earnings impacts, while welfare savings were spread evenly over the two lower-earnings subgroups. Similarly, applicants with a welfare history had most of the earnings gains and welfare savings, although both impacts were somewhat greater for the group with a briefer welfare stay.

Some characteristics associated with dependency and employability other than prior earnings and welfare history appear to be positively related to the program's impacts in San Diego. The results for subgroups presented in Table 4.6 suggest, for example, that race and the number of children in a household were important in this sample. These are clearly aspects of dependency, given the low earnings and the high welfare payments made to control group members who were non-white and had more than one child. Some of the other results are not consistent, notably the greater impacts for applicants who had a high school diploma or GED, a factor not usually related to long dependency. However, this may be due to the nature of the San Diego program. Prior education may have increased the probability of success in a program that (unlike Baltimore and Virginia) did not offer remedial education.

The subgroup results clearly indicate that the San Diego program had greater impacts on its most dependent applicants. Dependency, defined here as "having high welfare payments and low earnings," can be viewed as falling along a spectrum, ranging from the most to the least dependent cases, and as involving many characteristics, rather than just two. To assess the relationship between dependency, viewed this way, and program impacts, a "dependency score" was assigned to each person in the San Diego sample on the basis of a number of pre-program characteristics.[4]

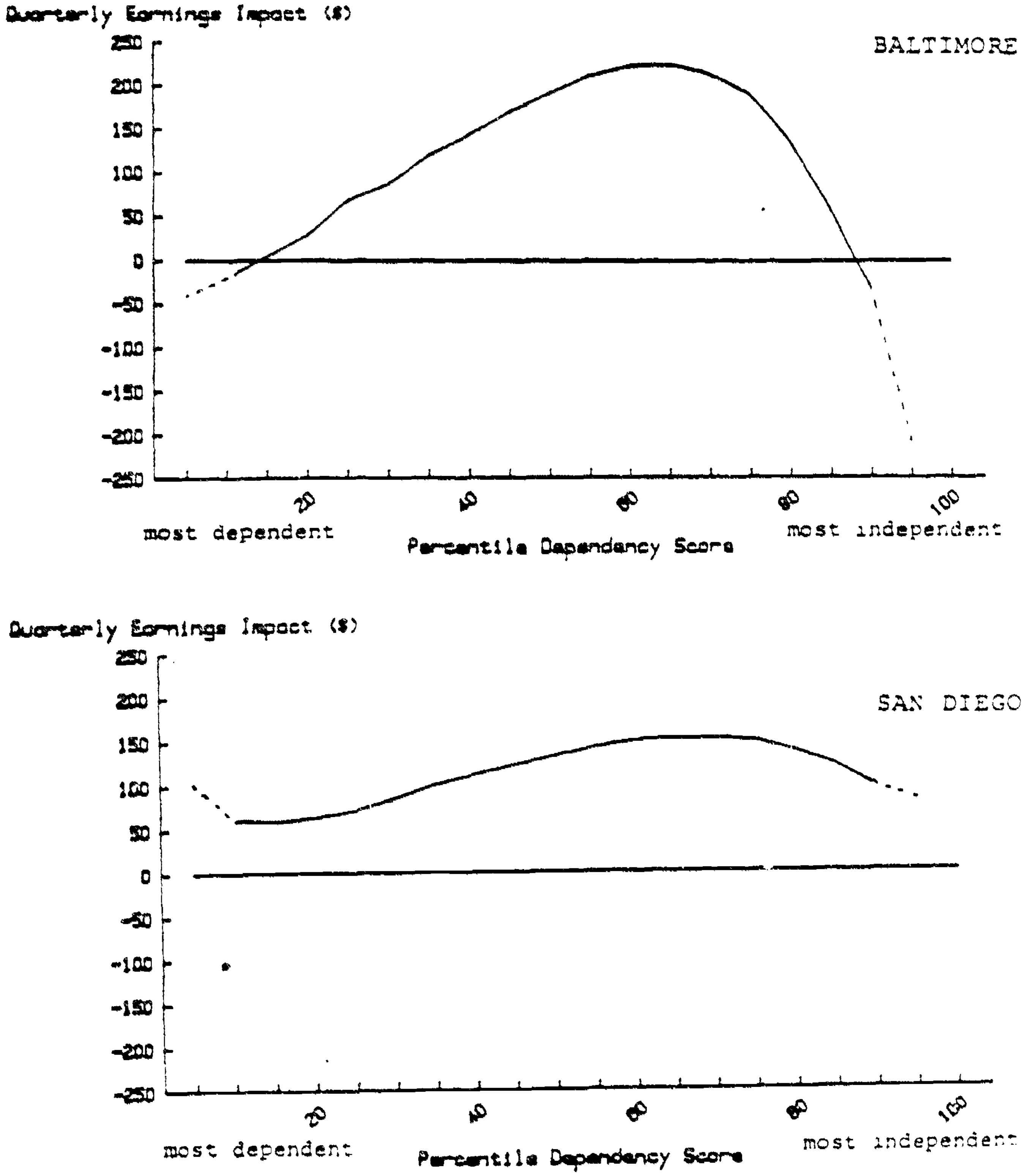
In Figure 4.1, earnings impacts estimated for individuals were plotted against their dependency scores. The bottom graph for the San Diego sample depicts the impact "responsiveness" of individuals with different levels of dependency. While the figure suggests that the San Diego program model was somewhat less effective with applicants at the two ends of the dependency spectrum, it also indicates that this or a similar program model was effective for a broad range of AFDC applicants. Even relatively dependent welfare applicants benefited somewhat from the treatment, which suggests that they should be included in such programs. And, while short-term job search and work experience may not always be helpful to relatively job-ready applicants, it was on average beneficial in San Diego.

The figure also shows a possible threshold effect: at some level of self-sufficiency and job-readiness, program impacts increased and, as seen in this graph, then began to decline again -- this time, for the least dependent in the sample. The Baltimore program, described below, is better suited to a discussion of this potential effect, since it enrolled a broad range of both applicants and recipients.

2. Baltimore

The Options program in Baltimore was very different from the San Diego initiative. Newly-mandatory AFDC recipients were enrolled as well as mandatory applicants. In addition, there was a greater range of services -- from independent job search to education and training -- and the services could vary according to the registrants' needs and preferences. Participant choice, however, was constrained by staff appraisal and slot availability. Because the least job-ready generally participated at higher

## FIGURE 4.1
## BALTIMORE AND SAN DIEGO EXPERIMENTALS: QUARTERLY EARNINGS IMPACTS, BY DEPENDENCY SCORE




NOTE: Dependency scores were estimated for San Diego and Baltimore experimentals based on their predicted earnings and predicted welfare receipt. The score is stated in percentile form, with "0" representing the most dependent. A score of 20 indicates that 20 percent of the sample ranked more dependent. Segments near endpoints of the curve are estimated with less precision and are therefore indicated with dashes.

levels in the more intensive services -- work experience, education and training -- than other cases, the subgroup impacts may have been influenced by the different services participants received, as well as their own characteristics.

The Baltimore results in this paper are based on an extra year of data collection, so the findings are somewhat different from those presented in the final report on the Options program. In the final report, it was suggested that the program treatment -- which could include services that lasted for a year or longer -- might lead to greater impacts at a later point. The additional data support that speculation. Overall earnings impacts continued to increase. Experimental-control differences in earnings for applicants were, in fact, more than 50 percent larger, and the earnings of recipients increased as well. Welfare savings, which earlier were not significant, did not change.

Two principal findings are apparent in the Baltimore subgroup impact results in Tables 4.4 and 4.5. One is the small earnings gains for recipients compared to applicants. As a whole, applicants in the program earned $172 more per quarter than controls, a statistically significant increase of 21 percent that is comparable to the change for applicants in San Diego. However, recipients, who were more welfare-dependent than applicants, earned only $37 more. The difference between applicants and recipients is statistically significant. These findings are especially important since recipients were not under-served, and the follow-up period was long enough to capture short-term program effects from education and training.

The pattern of subgroup results for applicants shows that those with

the highest pre-program earnings or without a welfare history had the smallest gains. The remaining large share of applicants did experience statistically significant earnings impacts. Among recipients, only those without pre-program employment experienced a statistically significant earnings increase.[5] No significant welfare savings were found in this longer follow-up period for either applicants or recipients or for any other subgroup.

Table 4.6 shows that Baltimore differed from San Diego in other subgroup earnings impacts. Applicants without a high school diploma or GED had larger impacts, perhaps reflecting the remedial education services offered by the Options program. Younger women and women with younger children also experienced somewhat larger-than-average gains. These factors operated differently in San Diego.

The top graph in Figure 4.1 plots the earning impacts of Baltimore applicants and recipients against individual dependency scores in the manner described for San Diego's applicants. The Baltimore program is particularly appropriate for such an investigation, since a broad spectrum of people, from first-time applicants to long-term recipients, were enrolled. The recipients, although all newly-mandatory, had often been on welfare in WIN-exempt status for some time. The combination of mandatory applicants and newly-mandatory recipients might be typical of an incoming group in a steady-state mandatory service program. However, the Baltimore program was limited to 1,000 slots to ensure adequate resources to serve the full range of enrollees.

The Baltimore graph, even more than in San Diego, lends substance to the threshold idea. It suggests that earnings impacts were largest for

individuals in the middle-dependency range, peaking at a point somewhat above the median. At the very left of the dependency spectrum, very dependent cases did not appear to respond to the services as well as other cases. Beyond some threshold level of self-sufficiency and job-readiness, program impacts increased. But at the other end of the spectrum, the program again had less effect -- this time on the people who were relatively well-prepared for jobs. These job-ready people seemed more able to enter employment and leave welfare without program help.

Two factors are important to note. The level of dependency was more extreme in the Baltimore sample because San Diego did not serve recipients. And, the shape of the curves for both programs would presumably change if the models or eligible populations changed.

3. Virginia

Virginia extended program participation requirements to the whole WIN-mandatory caseload of recipients as well as mandatory AFDC applicants. It also served rural as well as urban areas, and counties had considerable independence in implementing the program.[6] Resource constraints, however, were important: the program relied on job search assistance as its principal component and on independent job search as the most widely-used kind of job search. Community providers, such as schools and JTPA training programs, which received no program funding, provided the education and training. Because controls obtained these education and training services on their own with about equal frequency as the experimentals, the Virginia impacts can only be attributed to job search and work experience.[7]

Virginia has the shortest follow-up of the three programs -- only 4 to 6 quarters -- depending on the time an individual entered the sample. The

impact estimates are therefore preliminary, and should be interpreted with more caution than the others.

Statistically significant employment and earnings impacts were found only for applicants, not for recipients. As in Baltimore, the participation levels of recipients equalled or exceeded those of applicants, but recipient earnings impacts were about half those of applicants and not statistically significant. Welfare reductions were not statistically significant for either applicants or recipients.

Within the applicant group, sample members without recent pre-program employment experienced statistically significant increases in employment and reductions in the proportion on welfare. The highest prior-earnings group improved the least in both employment and welfare. However, the middle group -- based on prior earnings -- recorded the largest employment impacts. Applicants with no prior welfare had the smallest employment and welfare impacts. Earnings gains did not uniformly follow employment gains and, in fact, the impacts did not fall into the usual patterns seen for other high and low prior-earnings subgroups. These inconsistencies may stem in part from the limited follow-up data.

For recipients, the larger impacts were recorded for individuals with more welfare experience, although the prior-earnings categories did not exhibit much difference.

Results for the other subgroup categories in Virginia, which are presented in Table 4.6, show impact differentials among applicants of comparable magnitude to the differentials in the other states. The direction of these differentials, however, is not always the same. These variations in patterns across states suggest that different dimensions of dependency and

employability may dominate in particular program settings -- reflecting differences in programs, caseloads and labor markets.

C. Subgroup Combinations

One of the implications of the preceding analysis is that, not surprisingly, specific subgroup characteristics may differ in importance in different program settings. This implies that different dependency and/or employability criteria may be critical for a given program model or for a welfare caseload in certain locations. As a result, it is possible that subgroups defined in terms of many characteristics rather than just one may predict impact differences more consistently.

The combination of weak prior earnings with longer welfare history was used to define a more dependent portion of the sample. Table 4.7 presents impact results for four pairs of such subgroups. One pair shows applicants with prior earnings in the two lowest categories plus a welfare history versus those with either relatively high prior earnings or no welfare history. A similar split is made for recipients: no prior earnings and more than two years on welfare in one group against all other recipients in the second group. Two additional pairs were created by adding -- for the two more dependent groups -- the factor that group members did not have a high school diploma.

The results suggest that subgroups defined by combined work and welfare criteria may be more consistent predictors of impacts -- at least of earnings impacts -- than either characteristic alone. In all three applicant cases, the low-work, high-welfare subgroups experienced larger earnings impacts, although the differences for recipients were minimal.

TABLE 4.7

AFDC APPLICANTS AND RECIPIENTS: IMPACTS FOR SUBGROUPS
COMBINING PRIOR EARNINGS, PRIOR AFDC RECEIPT, AND HIGH SCHOOL DIPLOMA STATUS

| | Earnings Impact (Quarters 4 - Last) | | | | |
|---|---|---|---|---|---|
| | San Diego | Baltimore | | Virginia | |
| Subgroup | Applicants | Applicants | Recipients | Applicants | Recipients |
| Lower Prior Earnings Plus Higher Prior AFDC[a] | | | | | |
| No | + 83 | + 84 | + 3 | + 84 | + 38 |
| Yes | + 158** | + 254*** | + 50 | + 148** | + 58 |
| Lower Prior Earnings Plus Higher Prior AFDC Plus No High School Diploma | | | | | |
| No | + 108* | + 135* | + 22 | + 82 | + 46 |
| Yes | + 121 | + 254** | + 43 | + 169* | + 54 |

| | AFDC Payment Impact (Quarters 4 - Last) | | | | |
|---|---|---|---|---|---|
| | San Diego | Baltimore | | Virginia | |
| Subgroup | Applicants | Applicants | Recipients | Applicants | Recipients |
| Lower Prior Earnings Plus Higher Prior AFDC[a] | | | | | |
| No | - 4[illegible] | - 12 | + 2 | - 5 | + 0 |
| Yes | - 22 | - 15 | + 12 | - 41* | - 30 |
| Lower Prior Earnings Plus Higher Prior AFDC Plus No High School Diploma | | | | | |
| No | - 45* | - 6 | + 9 | - 16 | - 2 |
| Yes | + 10 | - 30 | + 5 | - 42 | - 42* |

SOURCE: See Table 4.1.

SAMPLE SIZE: Sample sizes are as follows: Applicants - San Diego = 2381, Baltimore = 1380; Virginia = 1269; and Recipients - Baltimore = 1377, Virginia = 1881.

NOTES: These data are regression-adjusted using ordinary least squares, controlling for pre-random assignment characteristics of sample members. Dollar-denominated estimates include zero values for sample members not employed and for sample members not receiving welfare.

[a]"Low prior earnings" is defined for applicants as earnings of less than $3000 in the year prior to random assignment; for recipients it is zero earnings. "Higher Prior AFDC" means any prior AFDC for applicants and more than two years for recipients.

A two-tailed t-test was applied to differences between experimental and control groups. Statistical significance levels are indicated as: * = 10 percent; ** = 5 percent; *** = 1 percent.

The addition of "lacks diploma" tended to reduce these differences.

## D. Program Costs

In developing welfare employment policy, program impacts on employment, earnings, welfare receipt and other outcomes must be weighed against program costs. This section briefly describes the cost differences by subgroup in the the San Diego and Baltimore programs and discusses the implications of these differences for the results of the overall analysis. A more detailed discussion of costs together with an assessment of the benefit-cost implications of the subgroup impact and cost differences in the benefit-cost analysis is available from MDRC.

Table 4.8 presents total program costs, expressed on a per experimental basis for the San Diego and Baltimore programs. The figures include the costs of serving nonparticipants as well as participants in the experimental groups, and are broken down by major program component. They are also disaggregated for the two major subgroups based on prior earnings and welfare experience.

Overall, subgroup variation in cost was small compared to the variation in impacts, particularly in San Diego, which had the same treatment sequence for all enrollees. Also, because that program was not long and education and training were not included in the sequence, costs were not large. The major components were group job search, work experience, assessment and support services. In Baltimore -- where total costs were higher -- relatively expensive services, such as education and training, were usually assigned to the less job-ready registrants. Thus, subgroup costs did vary somewhat, but the costs of services were closely related to

TABLE 4.8

PROGRAM COSTS PER EXPERIMENTAL, BY PROGRAM AND MAJOR SUBGROUP

| Subgroup | Total Average Cost | Group Job Search | Work Experience | Other Program Activities | Support Services |
|---|---|---|---|---|---|
| San Diego Applicants | $788 | $560 | $91 | $98 | $38 |
| Prior Year Earnings | | | | | |
| $3000 or More | 843 | 627** | 81 | 96 | 40 |
| $1-2999 | 733 | 522 | 81 | 96 | 35 |
| None | 775 | 537 | 103 | 96 | 39 |
| Had Own AFDC Case | | | | | |
| Never | 729 | 534 | 68*** | 96 | 33 |
| Two Years or Less | 794 | 567 | 91 | 96 | 40 |
| More Than Two Years | 845 | 585 | 124 | 96 | 41 |
| Baltimore Applicants[a] | 843 | 173 | 51 | 329 | 195 |
| Prior Year Earnings | | | | | |
| $3000 or More | 702* | 134 | 37* | 294 | 150* |
| $1-2999 | 949 | 204 | 50 | 376 | 218 |
| None | 879 | 183 | 63 | 323 | 214 |
| Had Own AFDC Case | | | | | |
| Never | 804*** | 166 | 28*** | 342*** | 177*** |
| Two Years or Less | 894 | 146 | 48 | 251 | 158 |
| More Than Two Years | 1037 | 208 | 70 | 408 | 247 |
| Baltimore Recipients[a] | 1065 | 188 | 89 | 386 | 288 |
| Prior Year Earnings | | | | | |
| $3000 or More | 831 | 192 | 55 | 344 | 159 |
| $1-2999 | 1041 | 215 | 82 | 363 | 267 |
| None | 1088 | 180 | 83 | 396 | 303 |
| Had Own AFDC Case | | | | | |
| Never | 835*** | 126 | 54* | 213* | 160** |
| Two Years or Less | 862 | 163 | 70 | 315 | 214 |
| More Than Two Years | 1156 | 200 | 97 | 420 | 320 |

SOURCE: MDRC calculations from program cost and enrollment data (see Long and Caspar, 1987).

NOTES: Estimates are total costs incurred for experimentals and are averaged over participants and non-participants. F-tests were performed on variation in cost in each column for each subgroup dimension. Statistical significance levels are indicated as: = 10 percent; ** = 5 percent; *** = 1 percent.

[a]The cost components listed for Baltimore do not include the costs of sanctioning, and thus do not sum to total cost.

specific groups -- for example, individuals without a high school diploma received more costly remedial education services than enrollees who already had a diploma.

Prior AFDC receipt was the most important characteristic associated with higher costs in both programs. The group with the longest stay on welfare -- more than two years -- had the highest costs. People in this subgroup stayed in the programs longer and, in Baltimore, were assigned to the expensive services more often.

The limited cost differences support the conclusions already reached. First, serving the less job-ready and the more welfare-dependent is cost-effective: while it costs slightly more to work with people who have been on welfare for a longer time, the net impacts on AFDC and employment are substantially larger than for the less dependent cases. Second, the less job-ready and the more dependent welfare applicants and recipients gain the most financially from these programs. Their earnings gains were generally more than enough to offset their reduced welfare benefits.

CHAPTER 5

MEASURES OF PROGRAM PERFORMANCE

As described in the first chapter, the best measure of performance in a welfare employment program is its impact on the people it serves. But genuine impacts cannot be obtained simply or quickly enough to be used in the management of most programs. This chapter assesses the value of two more practical measures of performance: first, "job entries" (including placements) and cases "off-welfare" (or welfare departures). It then discusses program participation and "coverage" indicators.

A. Job-Entry and Off-Welfare Measures

Program impacts require a comparison framework -- ideally, an experimental research design -- and data collection over a long enough period to fully observe the course of changes generated by a program. Unfortunately, much as administrators might want this information, they must make decisions in a reasonable period of time based on data that are readily available. Thus, the performance of welfare employment programs is usually determined by counting the numbers of registrants who obtain jobs and/or leave welfare. However, these measures overstate program impacts because, as the experience of the control groups in this analysis has shown, many recipients find jobs and leave welfare in the absence of program assistance. This means that some programs' high rates of job entry may result from their having relatively "job-ready" caseloads or a strong labor market, while the apparently poor performance of other programs may stem

from less advantageous conditions.

More serious than simple overstatement is the fact that, while change takes place when a recipient finds work or leaves welfare because of the program, the degree of that change varies by type of individual. A job entry for a recipient who has not worked for several years implies more change than a job entry by a person who has recently worked. This type of change -- or the degree of program success -- cannot be seen in unadjusted outcome measures. For example, conscientious program administrators seeking high job-entry rates may focus staff time and resources on placing relatively job-ready registrants, many of whom might have been able to find jobs on their own.

1. How Bad Are the Outcome Measures?

Using estimates of program impacts obtained on an individual basis are a logical way to assess the extent of the problem with outcome measures: the poorer the correlation between the performance measures and the impacts, the more serious the problem. Consequently, short-term job entry and off-welfare measures were examined in relation to program impacts on earnings and welfare payments in the San Diego and Baltimore programs.

A short-term job entry was defined as "employed at some point during quarters 2 or 3 after random assignment," and off-welfare was defined as "receiving no welfare payments in the third quarter." Somewhat longer-term measures took into account quarter 4 and the following ones for employment; quarter 6 was the point-in-time for welfare payments. In this report, the job entry data were UI earnings, which are more accurate and complete than typical program placement data,[1] but are less accessible to program operators.

Tables 5.1 and 5.2 display, in summary form, the results of correlating the job entry and off-welfare outcomes with program earnings and welfare impacts estimated for each experimental group member on the basis of regression results from the previous chapter.[2] The indicators are ranked in the table as "good" (positively correlated with impacts and statistically significant), "fair" (positively correlated but not statistically significant), "weak" (negatively correlated but not statistically significant) and "poor" (negatively correlated and statistically significant). Rankings are provided for all short-term versions of these indicators. If the longer-term version indicated substantial improvement, the higher rank is shown in brackets.

It is clear that job entries were not satisfactory performance measures for the San Diego or Baltimore programs. In all cases, short-term job entry was a weak or poor indicator of earning impacts, and the longer-term version showed little improvement. This suggests that the simple job entry measure of performance is inadequate and may even encourage reduced program performance in terms of impacts. Job entries were also not a satisfactory indicator of welfare savings.

The off-welfare measure also performed poorly. Most of the correlations with earnings gains and welfare savings were poor to fair. Interestingly, the off-welfare measure performed marginally better as an indicator of net welfare savings. One of the cases yielded a result of "good."

These findings are consistent with those of the previous chapter, supporting the conclusion that performance standards based on job entry or off-welfare rates are unrelated to true program effectiveness. The

TABLE 5.1

VALIDITY OF SIMPLE JOB ENTRY
PERFORMANCE INDICATOR

| Program and Welfare Status | Indicator Validity | |
|---|---|---|
| | Earnings Gain | Welfare Savings |
| San Diego, Applicants | poor | weak |
| Baltimore, Applicants | poor | poor [fair] |
| Baltimore, Recipients | poor | weak |
| Baltimore, All AFDC | weak | weak [fair] |

SOURCE: MDRC calculations from the County of San Diego welfare records, and Unemployment Insurance records from the EPP Information System; and from the State of Maryland welfare and Unemployment Insurance records.

NOTES: This table summarizes the correlations between the designated indicator and the earnings gains or welfare savings. The underlying numerical estimates are presented in an unpublished appendix. The following symbols are used:

Good indicates a correlation that has the correct sign and is statistically significant.

Fair indicates a correlation that has the correct sign but is not statistically significant.

Weak indicates a correlation that has the wrong sign but is not statistically significant.

Poor indicates a correlation that has the wrong sign and is statistically significant.

A longer-term version of the indicator was also tested in a second procedure by examining its partial correlation with predicted impact while controlling for the short-term indicator. If the partial correlation of a longer-term version raised the indicator's rank from the two lower to the two higher ratings, or from "fair" to "good," that change is noted in brackets in the table.

"Short- and longer-term" are defined as follows:

| | | |
|---|---|---|
| Short-term job entry | - | Any UI earnings in quarters 2 or 3 |
| Longer-term job entry | - | Any UI earnings in quarter 4 through last |
| Short-term off-welfare | - | No AFDC payments in quarter 3 |
| Longer-term off-welfare | - | No AFDC payments in quarter 6 |

TABLE 5.2

VALIDITY OF SIMPLE
OFF-WELFARE PERFORMANCE INDICATOR

| Program and Welfare Status | Indicator Validity | |
|---|---|---|
| | Earnings Gain | Welfare Savings |
| San Diego, Applicants | fair | weak |
| Baltimore, Applicants | poor | poor |
| Baltimore, Recipients | fair | good |
| Baltimore, All AFDC | fair | fair |

SOURCE AND NOTES: See Table 5.1.

findings should not be interpreted as suggesting it is wrong for programs to promote job entries or case closures. Rather, the results show that the outcomes of a program are closely tied to the characteristics of program registrants, and standards that ignore this fact provide a misleading picture of real program accomplishments. Such standards may also cause program funds to be poorly allocated.

2. Can Better Measures Be Developed?

Up to this point, a job entry has had equal value for all WIN-mandatory clients, regardless of their work and welfare histories. But the preceding chapter suggests a different scoring strategy -- one that gives more weight to job entries of registrants with weaker previous work records or longer time on welfare.

To explore this strategy, the job entry and off-welfare variables were calculated using a number of different weighting schemes indicating low earnings and high welfare dependency. Some of the indices were created on the basis of predicted levels of experimentals' earnings and welfare receipt. Others merely assigned extra points for job entries recorded for persons with low levels of pre-program earnings or with a long welfare history, using the definitions of the preceding chapters. The correlations improved in several cases, suggesting that the job entries of less employable welfare recipients should be given extra weight in setting performance standards.

Some of the tested weighting schemes used complex, regression-based indices. These methods require a complete demographic profile by enrollee and proper weights for each characteristic. While this approach may be suitable for aggregate analyses -- where proper weights can be calculated

for local labor market conditions and AFDC statutory grant levels -- they have drawbacks as a tool for local operators and caseworkers. The extra data collection is costly, and calculated scores for each enrollee would be subject to error. Perhaps most importantly, the complexity of the information may obscure the operational priorities line staff need.

The alternative approach uses information about only the most important registrant characteristics -- namely, prior employment and welfare experience. One such measure was created for job entries based only on prior employment:[3]

| | |
|---|---|
| Not employed in year prior: | 4 points per job entry |
| $1-2,999 earnings in year prior: | 2 points per job entry |
| $3,000 or more earnings in year prior: | 1 point per job entry |

The correlations of this weighted measure and welfare impacts are summarized in Table 5.3. A positive correlation between the indicator and the impact was found in all but one case, and the correlations were statistically significant in three instances. The longer-term version of the weighted indicator improved the results. Job entries were positively and significantly correlated with all earning impacts, and were also positively correlated to welfare impacts.

Job entries weighted this way were also positively correlated with the total net value of the program both to program registrants and to government budgets. These value estimates combine earnings and AFDC impacts, based on per-person estimates, as well as individual estimates of net program costs and program effects on taxes, Medicaid and other outcomes. Overall, then, job entries weighted by a person's prior earnings provide a simple-to-use performance measure that performed much better than

TABLE 5.3

VALIDITY OF WEIGHTED JOB ENTRY
PERFORMANCE INDICATOR

| | Indicator Validity | |
|---|---|---|
| Program and Welfare Status | Earnings Gain | Welfare Savings |
| San Diego, Applicants | good | good |
| Baltimore, Applicants | fair [good] | fair |
| Baltimore, Recipients | fair [good] | weak [fair] |
| Baltimore, All AFDC | good | fair |

SOURCE: See Table 5.1.

NOTES: This table summarizes the correlations between the designated indicator and the earnings gains or welfare savings. The underlying numerical estimates are presented in an unpublished appendix. The following symbols are used:

Good indicates a correlation that has the correct sign and is statistically significant.

Fair indicates a correlation that has the correct sign but is not statistically significant.

Weak indicates a correlation that has the wrong sign but is not statistically significant.

Poor indicates a correlation that has the wrong sign and is statistically significant.

A longer-term version of the indicator was also tested in a second procedure by examining its partial correlation with predicted impact while controlling for the short-term indicator. If the partial correlation of a longer-term version raised the indicator's rank from the two lower to the two higher ratings, or from "fair" to "good," that change is noted in brackets in the table.

"Short- and longer term" are defined as follows:

| | | |
|---|---|---|
| Short-term job entry | - | Any UI earnings in quarters 2 or 3 |
| Longer-term job entry | - | Any UI earnings in quarters 4 through last |
| Short-term off-welfare | - | No AFDC payments in quarter 3 |
| Longer-term off-welfare | - | No AFDC payments in quarter 6 |

Weights were assigned to job entry scores on the basis of prior earnings:

| | | |
|---|---|---|
| Not employed in year prior | - | 4 points per job entry |
| $1-2999 earnings in year prior | - | 2 points per job entry |
| $3000 or more earnings in year prior | - | 1 point per job entry |

unweighted job entries and case closure measures in the San Diego and Baltimore programs.

Nevertheless, this weighting scheme is not the final word on performance measurement. Employment-based measures may not be appropriate for programs with different goals (although the general principle of weighting could be adapted to other outcome measures, such as wage levels and job retention rates, which are not tested in this paper). Problems still remain in certain areas: i.e., determining the "points" to award job entries achieved without program assistance. Moreover, impacts alone may not provide the comprehensive picture of program participation sought by many administrators.

The next section briefly considers the implications of this research in developing some alternative performance measures of client activity in program components.

## B. Participation and Coverage

Performance measures based on program participation have often been used as an alternative, or an addition, to employment and welfare outcome measures. Compared to outcome measures, participation rates have both advantages and disadvantages. One clear advantage is that participation can be easily observed in the short term. One disadvantage is that the "intensity" of participation may not be easy to measure. For example, registrants in independent job search are counted as participants, but some of them have very little to do.

In-program activity measures have been important because many view participation as a precondition for impacts. However, such measures have

two problems. First, in mandatory programs, an "active participation" count ignores a good deal of program activity, much the same as a placement rate is a limited measure of program effects on employment. In mandatory programs, the behavior of nonparticipants is critical since nonparticipants may look for and find work or leave welfare in lieu of participating. Sanctioning and other program contact with nonparticipating individuals are explicitly intended to affect their behavior.

Second, participation may be less closely linked to impacts than short-term outcomes. Participation measures may cause staff to focus on the provision of services, whether or not individuals need them. A drive for high participation levels may result in program expenditures on those who are most likely to leave welfare on their own.

If participation measures are used, the subgroup impact findings indicate that priority should be given to registrants with poor work records. The same weighting scheme just applied to job entries can be used to develop weighted participation measures.

Another approach with considerable potential is the use of program "coverage" measures. Such measures have only been used in evaluation research, and have yet to be developed for use as program performance indicators. These measures would count, in addition to cases of participation *per se*, cases in which participation is no longer required or where sanctions for nonparticipation have been imposed. The concept of coverage takes into account the normal welfare caseload turnover, but it does so without requiring information about prior employment and welfare and need not involve weights.

Under a coverage formula, a client might be counted as "covered" by

program requirements if any of these outcomes is achieved:

1. Completes or is completing program requirements;
2. Becomes employed;
3. Leaves AFDC; or
4. Is sanctioned for nonparticipation.

To maximize coverage, the focus of administrators is automatically directed to the longer-term recipients, who are more likely to remain uncovered. People who seem likely to remain on welfare and enrolled in the program will receive attention. This differs from programs in which unweighted participation measures are used, where the participation of short-term welfare recipients "counts," even if they would have left welfare quickly without special services. With coverage measures, programs have less incentive to serve only the most job-ready enrollees, since a client can be counted only once as covered either through "participation" or "placement."

Data for experimentals in the three programs studied illustrate how a coverage measure might work in practice. In these programs studied, only from 10 to 20 percent of experimentals were still on welfare nine months after enrollment and had not begun employment, had not participated in any major component, or had not been sanctioned for not participating. (At any point in time during the nine-month period, however, the coverage rate would be lower.) This conveys a useful overall impression to legislators and the public about how a program is managing to work with its eligible caseload. In addition, because some two-thirds of the "non-covered" experimentals in the studied programs were recipients, and three-quarters of this group had no prior earnings, a coverage standard for welfare employment programs could shift attention toward these more dependent subgroups.

No short-term performance indicator is ideal. However, this analysis indicates that, in welfare employment programs, measures should take account of differences in the welfare dependency and employability of the individuals served. In principle, any of several indicators, combined judiciously with other information, can be used to measure program performance. This analysis suggests that weighted outcome measures correct some of the defects of common unweighted measures. Coverage measures also hold promise. The second phase of the research will provide additional information on choosing appropriate program measures and standards.

## FOOTNOTES

## CHAPTER 1

1. Results of the full benefit-cost analysis are described in an internal working paper by the authors. The findings of this analysis generally parallel the impact findings.

2. The use of the term "placement" is avoided in this paper. The term was originally used by the employment service to denote referral of a client to a particular job opening by program staff. It is therefore inappropriate for programs that rely on a client's own job search efforts. In addition, placements, or self-reported employment, tend to understate employment and earnings because recipients sometimes do not report jobs to welfare staff.

   Similarly, the term "off-welfare" is used rather than "case closure" because it is more inclusive. It covers persons who apply for AFDC, enter a program, but then quickly leave the welfare system without having been approved for a grant (i.e., without ever having had a case opened).

   "Off-welfare" and "welfare reduction" indicators are not identical. The former looks only at whether families are receiving any AFDC payment, and it is stated as a numerical count or as a percent. The various welfare reduction formulas in use subtract pre-program welfare grant levels for clients from their post-program welfare receipt to arrive at a dollar figure, either aggregate or per registrant. The first phase of this study tests an off-welfare indicator rather than a welfare dollar reduction indicator because the pre-program data necessary to simulate that indicator is lacking from the San Diego and Baltimore research data bases.

3. The role of performance scores in the actual distribution of funds has been quite small. The bulk of federal WIN funds have been allocated to states according to number of WIN registrants. On the basis of budget appropriations during the 1970s, it has been determined that incentive rewards for performance based on this formula could amount to about one-third of all federal WIN moneys given to states. (See Office of Family Assistance, 1985, pp. 13-14.)

   In practice, annual funding changes have been restricted in other ways. WIN regional coordinators have had discretionary powers, and incentive moneys could be allocated for local performance achievements not incorporated in the mathematical formula or on the basis of other considerations. As a result,

only about 3 percent of funds distributed in a given year have reflected performance scores, although cumulative changes across the years could have amounted to more. (Ibid, p. 21.)

Job retention has been a more important determinant of the program performance score in the discretionary part of the WIN Allocation Formula than job entry, although there is some evidence that the complexity of the formula kept this fact hidden from line operators (Mitchell, Chadwin, Nightingale, 1980, p. 287). The relative potential of each element of the formula to raise a state's overall performance score differed, depending on how high or low its score on each element might be. The complexity of the discretionary part of the formula was such that determining which elements had the greatest influence on scores would be very difficult without sophisticated analysis and simulation.

4. Participation is observed now, whereas outcomes may be observed only after some months and may require substantial effort in locating clients to ask about their employment status. Monitoring subgroup participation may be the most effective way of ensuring local compliance with an optimal targeting plan.

5. The problem of specifying optimal performance standards for independent local service providers for JTPA programs has been highlighted by the growing use of fixed-priced contracting. The language of JTPA has encouraged the use of fixed-priced contracting because all costs incurred can be allocated to "training," thus helping programs to comply with the 15 percent cap on administration costs. For a thorough discussion of the possibilities and problems in fixed-priced contracting see Wallace, 1985.

6. Indicators that make use of pre-program client characteristics are often referred to as change-based indicators, with simple outcomes designated as level indicators. The example given in this chapter for San Diego would suggest that change- based indicators should prove superior to simple outcomes as proxies for real program impact. In that case, the change from no pre-program employment to employment during the follow-up period was associated with the larger program-induced impact on employment. The weighted job entry rates tested in this paper are change-based indicators, since they award more performance points for the employment of clients who were not employed in the recent pre-program period.

The relevant literature on indicator validation is based on several analyses of CETA. Borus, 1978, found that job entry had very little power to indicate net impact for CETA. Gay and Borus, 1980, in a study of four pre-CETA programs, found

change indicators to be somewhat superior, and rated simple job entry as one of the poorest measures. In contrast, Geraci and King, 1981, found evidence supporting job entry as the better measure, as did Geraci, 1984. Zornitsky et al., 1985, produced results favoring level indicators. The latter three studies also concluded that post-program follow-up added valuable information about employment at the point of termination.

These studies all suffer serious methodological problems from having been based on non-experimental impact estimates. The principal issue -- the value of level indicators versus change-based indicators -- is still the most pressing one to be resolved in performance monitoring. The issue is complicated by the possiblity that the best class of indicators may be different for welfare women, adult men and youths. Adult men entering employment programs typically exhibit a temporary pre-program dip in earnings, making prior earnings problematic as a proxy for earnings capability. Youth often have short and erratic earnings histories, and a pre-program earnings baseline may therefore be meaningless for them.

7. See Bane and Ellwood, 1983; Ellwood, 1986.

8. See Ellwood, 1986, p. xii.

9. The wait-and-see approach does not rely on an ability to predict future dependency and does not face the political hurdle of denying services to subgroups based on marital status and age of youngest child. On the other hand, an initial period may have been wasted, a period in which improvements could have been made. Ellwood in his 1986 work suggests that evidence favors early identification and targeting over the wait-and-see strategy.

10. See O'Neill et al., 1984, p.84.

11. See MDRC, 1980.

## CHAPTER 2

1. See Goldman et al., 1986; Friedlander et al., 1985; and Riccio et al., 1986. For a summary of the demonstration's findings thus far, see Gueron, 1987.

2. In San Diego, a second experimental group received job search only. The program and its evaluation were also carried out for AFDC-U. Neither of these research groups is analyzed in this study.

3. In this report, participation and sanctioning rates were calculated on somewhat different bases than in the published state reports. In this study, the base is always "all experimentals." In the state reports, the base of "all program registrants" was often used. Most experimentals did, however, register for the programs, and the differences between the figures cited here and those published in the state reports are not large.

4. Sample sizes in this report differ slightly from those in the corresponding state reports. An attempt was made here to assign values to demographic data where these were missing. If missing data could not be inferred with reasonable certainty, the cases were dropped from the analysis. The effect on sample size was the gain of 7 cases in San Diego and 54 cases in Baltimore, but a loss of 32 cases in Virginia.

5. Randomization produced similar experimental and control groups with, however, some differences. There were small differences between research groups in ethnicity and marital status in the San Diego sample. In the other two samples, small differences were apparent in measures of education, prior employment and earnings.

6. This does not mean that the indicated subgroups account for the bulk of <u>all</u> AFDC expenditures. Benefits paid to families outside of the WIN-manadatory sample are not counted. Nationally, about two-thirds of AFDC families are WIN-exempt.

## CHAPTER 3

1. For more complete reports of data quality control, see the individual state reports.

2. For more detail about data sources and follow-up, consult the state reports.

3. The distinction between unconditional and conditional impact estimates can be developed as follows. The basic impact regression model is

   Y(T, S1, S2, X)

   where

   | | |
   |---|---|
   | Y | outcome variable |
   | T | experimental group dummy variable |
   | S1 | dummy variable for subgroup dimension 1 |

S2 dummy variable for subgroup dimension 2

X vector of additonal control variables

The full sample impact is the coefficient of T. The unconditional subgroup estimates for S1 come from the regression model

Y(TS1, TNS1, S1, S2, X)

where

TS1 = T * S1

TNS1 = T * (1-S1)

The impact on groups S1=1 and S1=0 are read from the coefficients of TS1 and TNS1, respectively. Finally, the conditional model is

Y(T, TS1, TS2, S1, S2, X)

where

TS2 = T * S2

and the coefficient of T is the impact when S1=0 and S2=0. The coefficient of TS1 is the additional impact attributable to the S1 characterstic when S2 is held constant. The coefficient of TS2 is the additional impact attributable to the S2 characteristic when S1 is held constant.

Interactive specifications are possible for both unconditional and conditional models. For the unconditional case,

Y(TS12, TS1N2, TSN12, TSN1N2, S1, S2, S12, X)

where

TS12 = T * S1 * S2

TS1N2 = T * S1 * (1-S2)

TSN12 = T * (1-S1) * S2

TSN1N2 = T * (1-S1) * (1-S2)

S12 = S1 * S2

For the conditional case,

Y(T, TS1, TS2, TS12, S1, S2, S12, X)

Coefficients in this latter model can be combined to reproduce the unconditional interaction estimates exactly. But when a third subgroup dimension is introduced, S3, the term TS3 in the conditional model would make the two sets of interaction estimates different.

4. See Borus, 1978.

5. Individual impact estimates are made by (1) regressing demographic and background characteristics on employment and welfare outcomes for the experimental and control groups, and then (2) using the coefficients obtained from these regressions, along with the characteristics of individual members of the experimental group, to predict individual impacts. The first stage estimate is made from the conditional subgroup impact regression model. That is, from the regression that contains the full array of experimental subgroup interactions, a prediction is made forr the expected program impact on earnings and welfare receipt for each person in the experimental sample. The net impact estimate will differ for each person, depending on the demographic, and prior work and welfare characteristics at the time of entry into the research sample.

   These are sometimes referred to as direct estimates. For example, with treatment interactions for prior employment, education and number of children, one impact would be predicted for an experimental with no prior employment, no diploma, one child; a different net impact would be predicted for an experimental with any difference in any of these characteristics. The more variance in the dependent variable that can be accounted for by the regression model, the better the predicted net impacts. At the present state of knowledge, however, most of the variation in the outcome measures cannot be explained.

## CHAPTER 4

1. The applicant/recipient distinction is often a significant one for program operators, as it was in San Diego. Also, F-tests for homogeneity of regression coefficients have consistently turned up large differences in regression models for applicants and recipients in welfare receipt equations. For these reasons, and to more easily handle expected differences in applicant and recipient behavior, the samples were split for the regression runs.

2. The composite estimates are a weighted average of estimates of the impacts and adjusted means for the five estimation subsamples. The weights are the inverse estimated standard error for each impact estimate, normalized by dividing by the sum of the inverse standard errors. The choice of weights minimizes the variance of the composite estimate, satisfying one of the objectives of pooling. Another choice of weights could have been the fraction of the total sample accounted for by each of the five estimation samples. But the designs in San Diego and Virginia are unbalanced, with about a 2:1 experimental-control ratio, and the interpretation of such a weighting scheme is not clear. A final alternative would have been to weight each sample by the fraction of all work/welfare program enrollees in the country who are in programs similar to each of the three under study here, an endeavor beyond the scope of this paper.

3. For this analysis, impact regressions were run on the pooled sample of applicants and recipients, first in Baltimore and then in Virginia. The model specified an experimental group dummy, a dummy for applicants, and a dummy for an experimental-applicant interaction. This last dummy gave the estimate of the unconditional impact difference. Interactions of experimental group membership with all other subgroup characteristics were then added and the same coefficient read again. The t-statistic for this coefficient therefore gives the statistical significance of the conditional difference in impacts between applicants and recipients. Applicant/recipient differences in earnings gains were statistically significant in Baltimore but not in Virginia.

4. A dependency index was created as follows. Average earnings and average AFDC dollars received were regressed on demographic variables for control group clients in Virginia. These coefficients were then used to predict follow-up earnings and welfare benefits for sample members in San Diego. The index variable was created as predicted earnings minus predicted welfare. An earnings impact regression was then run for San Diego using linear through quartic terms in the index and linear through quartic terms in the interaction of the experimental group dummy with the index, plus the experimental group dummy itself. This dummy and the four interaction coefficients were then used to plot predicted impacts at 5-percent n-tile points of the index variable. The procedure was repeated for Baltimore.

5. The negative earnings impacts for the subgroup with some year-prior earnings may indicate that the longer-term employability activities for welfare recipients with an employment record keep such persons out of the labor market when they would have been working. It seems likely, however, that a major part of

the negative differential is anomalous, a product of chance. The recipient control group in this prior earnings category had higher earnings than the corresponding applicant controls, whereas <u>all</u> the other recipient control subgroups earn less than their applicant control counterparts. This suggests that the true earnings losses of this recipient subgroup might not be as severe if the experiment were to be replicated.

6. AFDC benefit levels also vary across counties in Virginia.

7. See Riccio et al., 1986, p. xiv.

## CHAPTER 5

1. Under-reporting of job entries can occur when case heads who leave welfare because they have found jobs do not report employment. Particularly in large urban areas with large caseloads, cases are often closed because the client fails to respond to some attempt at contact, making it impossible to record employment status or other eligibility factors. In addition, reports of employment obtained by income maintenance staff for the purpose of adjusting grant payments are not always reported back to the staff of the employment program.

2. Regressions for average earnings and average welfare payments over quarter 4 through the last quarter were run with all treatment-subgroup interactions in the model at once. The coefficients of these interactions were then used to predict for every experimental group member the expected net impact on earnings and welfare receipt. These new variables were then correlated with employment and off-welfare status, using only the experimental group sample.

3. These weights represent approximately the relationship of control group mean earnings for prior-earnings categories in the composite impact table in the preceding chapter.

REFERENCES

Bane, Mary Jo; and Ellwood, David T. 1983. "The Dynamics of Dependence: The Routes to Self-Sufficiency." Cambridge, Mass.: Urban Systems Research and Engineering, Inc.

Borus, Michael E. October 1978. "Indicators of CETA Performance." Industrial and Labor Relations Review, Vol. 32, No. 1.

Ellwood, David T. January 1986. "Targeting 'Would-Be' Long-Term Recipients of AFDC." Princeton, N.J.: Mathematica Policy Research, Inc.

Friedlander, Daniel; Hoerz, Gregory; Long, David; and Quint, Janet. 1986. Maryland: Final Report on the Employment Initiatives Evaluation. New York: Manpower Demonstration Research Corporation.

Gay, Robert S.; and Borus, Michael E. Winter 1980. "Validating Performance Indicators for Employment and Training Programs." The Journal of Human Resources, Vol. 15, No. 1.

Geraci, Vincent J. August 1984. "Short-Term Indicators of Job Training Program Effects on Long-Term Participant Earnings." University of Texas-Austin, Center for the Study of Human Resources, Project Working Paper 2.

Geraci, Vincent J.; and King, Christopher T. September 1981. "Employment and Training (CETA) Program Performance: Long-Term Earnings Effects and Short-Term Indicators." University of Texas-Austin, Center for the Study of Human Resources, Working Paper Series Number 40-81.

Goldman, Barbara S. 1981. Impacts of the Immedidate Job Search Assistance Experiment: Louisville WIN Research Laboratory Project. New York: Manpower Demonstration Research Corporation.

Goldman, Barbara; Friedlander, Daniel; and Long, David. 1986. California: Final Report on the San Diego Job Search and Work Experience Demonstration. New York: Manpower Demonstration Research Corporation.

Grossman, Jean Baldwin; Maynard, Rebecca; and Roberts, Judith. October 1985. Reanalysis of the Effects of Selected Employment and Training Programs for Welfare Recipients. Princeton, N.J.: Mathematica Policy Research, Inc.

Grossman, Jean Baldwin; and Mirsky, Audrey. February 1985. A Survey of Recent Programs Designed to Reduce Long-Term Welfare Dependency. Princeton, N.J.: Mathematica Policy Research, Inc.

Gueron, Judith. 1987. Reforming Welfare with Work. New York: The Ford Foundation.

Manpower Demonstration Research Corporation. 1980. Summary and Findings of the National Supported Work Demonstration. Cambridge, Mass.: Ballinger Publishing Company.

Mitchell, John J.; Chadwin, Mark L.; and Nightingale, Demetra S. 1980. Implementing Welfare-Employment Programs: An Institutional Analysis of the Work Incentive (WIN) Program. Washington, D.C.: U.S. Department of Labor, Employment and Training Administration, Monograph No. 78.

Office of Family Assistance, Department of Health and Human Services. April 1985. "A Study of the Work Incentives (WIN) Allocation Formula." Washington, D.C.: Office of Family Assistance.

O'Neill, June A.; Wolf, Douglas A.; Bassi, Lauri J.; and Hannan, Michael T. June 1984. An Analysis of Time on Welfare. Washington, D.C.: The Urban Institute.

Riccio, James; Cave, George; Freedman, Stephen; and Price, Marilyn. 1986. Virginia: Final Report on the Employment Services Program. New York: Manpower Demonstration Research Corporation.

Wallace, John W. January 1985. "Information Collection for the Job Training Partnership Act: Performance Standards vs. Evaluation." Washington, D.C.: National Commission for Employment Policy.

Wolfhagen, Carl. 1983. Job Search Strategies: Lessons from the Louisville WIN Laboratory. New York: Manpower Demonstration Research Corp.

Zornitsky, J.; Schneider, G.; Sharick, R.; and Shapiro, E. June 1985. Post-Program Performance Measures for Adult Programs Funded Under Titles IIA and III of JTPA: Early Findings Prepared for the U.S. Department of Labor's Performance Standards Advisory Committee. Cambridge, Mass.: Abt Associates, Inc.